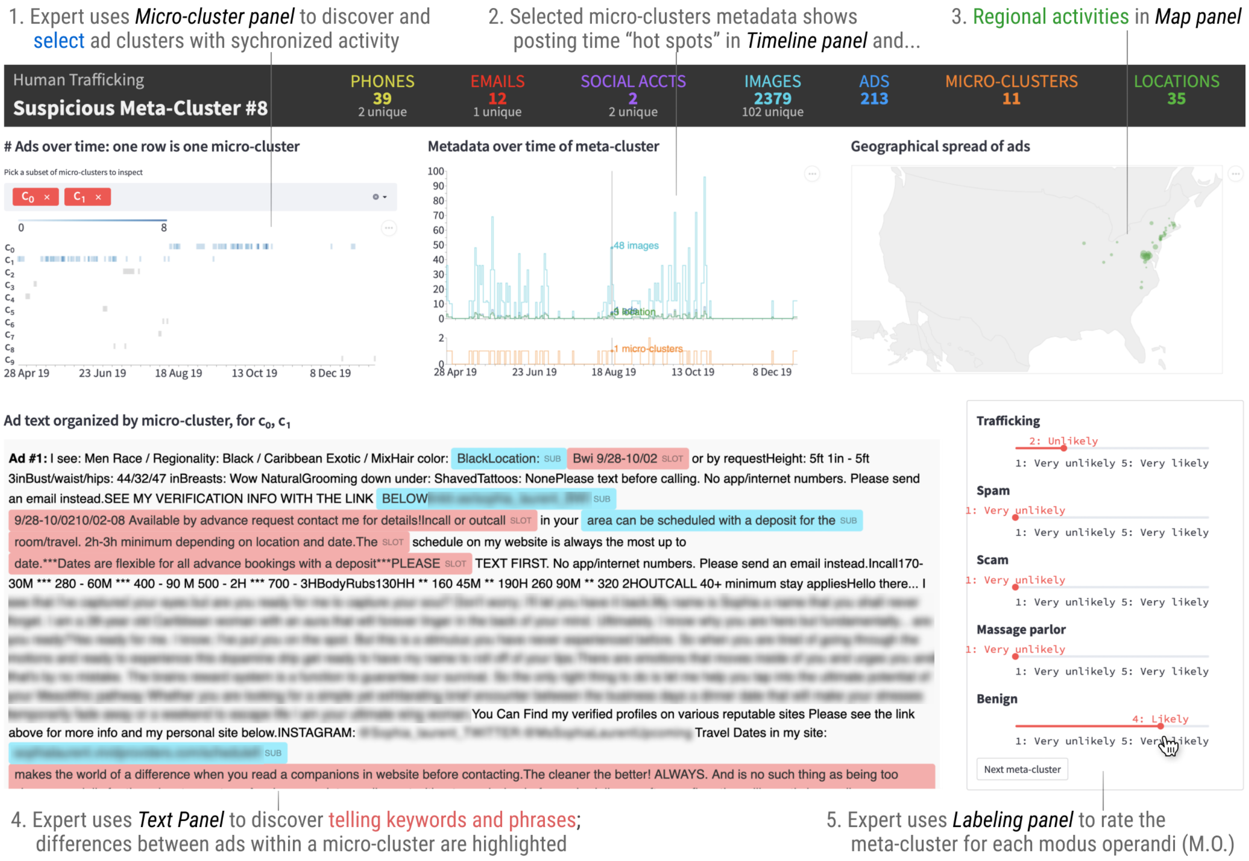
TRAFFICVIS: Visualizing Organized Activity and Spatio-Temporal Patterns for Detecting and Labeling Human Trafficking

Catalina Vajiac[*] Duen Horng Chau[†] Andreas Olligschlaeger[‡] Rebecca Mackenzie[§]
Pratheeksha Nair[¶] Meng-Chieh Lee[*] Yifei Li[¶] Namyong Park[*] Reihaneh Rabbany[¶]
Christos Faloutsos[*]



1. Expert uses *Micro-cluster panel* to discover and select ad clusters with sychronized activity

2. Selected micro-clusters metadata shows posting time "hot spots" in *Timeline panel* and...

3. Regional activities in *Map panel*

4. Expert uses *Text Panel* to discover telling keywords and phrases; differences between ads within a micro-cluster are highlighted

5. Expert uses *Labeling panel* to rate the meta-cluster for each modus operandi (M.O.)

Ad #1: I see: Men Race / Regionality: Black / Caribbean Exotic / MixHair color: BlackLocation: SUB Bwi 9/28-10/02 SLOT or by requestHeight: 5ft 1in - 5ft 3inBust/waist/hips: 44/32/47 inBreasts: Wow NaturalGrooming down under: ShavedTattoos: NonePlease text before calling. No app/internet numbers. Please send an email instead.SEE MY VERIFICATION INFO WITH THE LINK BELOW[redacted] SUB 9/28-10/0210/02-08 Available by advance request contact me for details!Incall or outcall SLOT in your area can be scheduled with a deposit for the SUB room/travel. 2h-3h minimum depending on location and date.The SLOT schedule on my website is always the most up to date.***Dates are flexible for all advance bookings with a deposit***PLEASE SLOT TEXT FIRST. No app/internet numbers. Please send an email instead.Incall170-30M *** 280 - 60M *** 400 - 90 M 500 - 2H *** 700 - 3HBodyRubs130HH ** 160 45M ** 190H 260 90M ** 320 2HOUTCALL 40+ minimum stay appliesHello there... I [redacted] You Can Find my verified profiles on various reputable sites Please see the link above for more info and my personal site below.INSTAGRAM: [redacted] Travel Dates in my site: [redacted] SUB
makes the world of a difference when you read a companions in website before contacting.The cleaner the better! ALWAYS. And is no such thing as being too

Trafficking
2: Unlikely
1: Very unlikely 5: Very likely

Spam
1: Very unlikely
1: Very unlikely 5: Very likely

Scam
1: Very unlikely
1: Very unlikely 5: Very likely

Massage parlor
1: Very unlikely
1: Very unlikely 5: Very likely

Benign
4: Likely
1: Very unlikely 5: Very likely

Fig. 1. **Analyzing online escort ads using TRAFFICVIS:** we show one meta-cluster, i.e. micro (text) clusters connected using metadata, on real data. Some text blurred for privacy. **1.** Human trafficking domain expert uses *Micro-cluster panel* to drill down to specific micro-cluster data and associated ads. **2-3.** Expert uses *Timeline panel* and *Map panel* to investigate metadata, noticing inconsistent posting time and regional geographic spread, ruling out spam and scam. **4.** Expert uses *Text panel* to quickly find telling signals; differences between ads in a micro-cluster are highlighted. **5.** Finally, the expert confidently labels the meta-cluster for each modus operandi (M.O.), deciding on *benign* (at-will sex worker), with a small chance of *trafficking*.

Abstract— Law enforcement and domain experts can detect human trafficking (HT) in online escort websites by analyzing suspicious clusters of connected ads. How can we explain clustering results intuitively and interactively, visualizing potential evidence for experts to analyze? We present TRAFFICVIS, the first interface for cluster-level HT detection and labeling. Developed through months of participatory design with domain experts, TRAFFICVIS provides coordinated views in conjunction with carefully chosen backend algorithms to effectively show spatio-temporal and text patterns to a wide variety of anti-HT stakeholders. We build upon state-of-the-art text clustering algorithms by incorporating shared metadata as a signal of connected and possibly suspicious activity, then visualize the results. Domain experts can use TRAFFICVIS to label clusters as HT, or other, suspicious, but non-HT activity such as *spam* and *scam*, quickly creating labeled datasets to enable further HT research. Through domain expert feedback and a usage scenario, we demonstrate TRAFFICVIS's efficacy. The feedback was overwhelmingly positive, with repeated high praises for the usability and explainability of our tool, the latter being vital for indicting possible criminals.

Index Terms—Human trafficking, Labeling, Visualization, Infoshield.

[*] Carnegie Mellon University
[†] Georgia Institute of Technology
[‡] Marinus Analytics
[§] University of Pittsburgh
[¶] McGill MILA

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication

1 INTRODUCTION

Human trafficking (HT) for forced sexual exploitation is a pervasive societal problem that affects over 4.8 million people world-wide [22], and the majority of HT victims are advertised on online escort websites [31,37]. However, legitimate sex workers also post on these sites, and law enforcement agencies are focused on separating HT rings from legitimate sex workers. There is one critical insight to detecting HT; since traffickers entirely control the ad content for their victims [33,37], ads posted by the same trafficker tend to be similar.

However, the problem is more complex; domain experts have recently discovered additional modus operandi (M.O.s) in escort ads. For example, *spam* ads with fake contact information flood escort websites, and *scam* ads ask for prepayment and don't provide any services. These M.O.s have made HT detection more difficult for law enforcement. *Spam* ads are particularly problematic – there are often very many of them posted in a very short timespan, flooding escort websites and often looking like leads at first glance. Features of these M.O.s are not yet well known, as it is currently prohibitively time-consuming for domain experts to label clusters so that researchers can analyze them. Currently, the only cluster labels we have are from domain experts stumbling upon them haphazardly; in fact, we do not have even a single labeled cluster for some M.O.s. However, having good cluster labels would enable the development and evaluation of M.O. classification algorithms.

Furthermore, once these algorithms are developed, they need to be accessible to all domain experts – many of which do not have a strong background in AI. Law enforcement officers and prosecuting attorneys are among the expected users of these algorithms.

Given the text, contact information, and posting times of each advertisement (images are not analyzed due to the pervasiveness of trafficked minors in this data), how can we best find groups of ads that are indicative of HT? Current methods for fighting HT [28] find text-clusters (referred to as *micro-clusters*). However, these methods cannot help domain experts find possibly suspicious clusters if there is no intuitive way for them to interact with the results. Developing useful visualizations for HT results is challenging due to the multimodal aspect of the data; clusters involve large amounts of text, spatio-temporal posting behaviors and metadata, all of which contain insights that influence final labeling decisions.

We propose TRAFFICVIS, an interactive application for domain experts to visually inspect suspicious meta-clusters (micro-clusters connected with metadata) and label their likelihood to be a particular M.O. Developed through months of participatory design with domain experts, TRAFFICVIS provides coordinated views in conjunction with carefully chosen backend algorithms to effectively show spatio-temporal and text patterns to a wide variety of anti-HT stakeholders. We extend the ideas presented in [48] by extending our participatory design with domain experts and evaluating TRAFFICVIS's efficacy with solicited domain expert feedback. In particular, TRAFFICVIS makes the following contributions:

1. **High-impact:** TRAFFICVIS is the first interactive application for HT detection and understanding. Experts say TRAFFICVIS is accessible to a variety of anti-HT stakeholders, including criminologists, domain experts, and law enforcement (see Section 6).

2. **Label generation:** Curating human-generated labels is a notoriously difficult and time-consuming process. TRAFFICVIS allows domain experts to label hundreds or thousands of ads at once, enabling researchers to develop and evaluate better HT detection algorithms down the line.

3. **Time-saving:** Through expert feedback, we find that it takes between 2-4 minutes for an expert to label clusters using TRAFFICVIS. Experts estimate it would take at least 20-30 minutes to investigate clusters with any other method (see Section 6). This provides a vast speedup vs the current standard, manual labeling, finally enabling domain experts and law enforcement to sift through these ads efficiently.

Reproducibility: The code is open-sourced at `https://github.com/catvajiac/TrafficVis`. We also provide synthetic data to demonstrate how TRAFFICVIS might be adopted while guarding against privacy risks, ensuring others can try our tool without compromising victims' safety.

2 BACKGROUND AND RELATED WORK

We first discuss previous work related to HT, then label generation in the machine learning and visualization communities.

2.1 Existing work on HT.

To our knowledge, no published work exists for HT visualization, but there are some known algorithms for HT detection.

Industry solutions. SpotLight [44] and Marinus [30] focus on fighting HT. Their existence shows that law enforcement is interested in software solutions for HT detection. While both have made advances, they don't visualize clusters of related ads and text simultaneously. As far as we know, none use text processing to connect related ads. Furthermore, these solutions are proprietary, which limits their reach.

Advertisement level solutions. A few HT detection methods focused on advertisement level classification, rather than a clustering task [3,18,25,46]. Many of these methods relied on specific indicators to mark ads as suspicious, such as keywords indicating underage victims. However, due to the adversarial nature of HT, predefined features will not stay relevant over time. Supervised methods used text and image data to predict the suspiciousness of an ad [46, 50] on a particular dataset. Unfortunately, the dataset used in these algorithms has noisy and biased labels; ads containing the word "Asian" are significantly more likely to be flagged as HT, irrespective of whether they actually were or not. Also, these ads are old, and posting behavior has dramatically changed since then, especially given the fall of Backpage [42], which greatly disrupted the escort market. Most problematically, these methods cannot find *groups* of organized activity, which is problematic for law enforcement – if a particular trafficker is being investigated, they need to discover all ads that are relevant to understand the scope of HT and quickly provide help to victims.

Cluster level solutions. Some related work tries to find connections between ads. Some methods train binary classifiers to predict if two ads are connected [32,38], while others uses local active search approach to retrieve connected ads [39]. Sentence-level embedding and hashing techniques have also been used to find groups of ads [29]. InfoShield [28] uses Minimum Description Length to create these templates, containing no black-box components.

Unfortunately, none of these published methods include interactive visualizations. Furthermore, even the cluster level solutions make the assumption that each text-based cluster represents a different organized activity. However, this is not always the case – many traffickers change the templates they use to write ads over time, or use different templates in different regions. We aim to exploit the connections between these text-based clusters in our visualization. In this paper, we will build our system upon the micro-clusters (text clusters) found by InfoShield.

2.2 Label generation systems.

Labeling is a crucial step to the development and evaluation of machine learning models on real-world data, but is also notoriously a labor and time-intensive process [8,51].Often, particularly for simple labeling tasks, crowdsourcing marketplaces such as Amazon Mechanical Turk (MTurk) are used to quickly generate labels [11]. However, since the profiles of workers are not well known, there are quality issues associated with MTurk [19,23]. Furthermore, it is not well-suited to problems where significant domain knowledge is required.

Broadly, visualization can provide valuable knowledge to the analyst [13,49], which can help the labeling process [9]. These labels are commonly used in two ways: either in real time to improve the

xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

performance of the current task, such as in active learning settings, or for the collection of data to be used in downstream tasks.

Active learning approaches. There is much work on active learning and the role of visualization in improving these algorithms [12]. Some systems focus on recommending the most probable labels based on semi-supervised models on a larger set of disjoint labels [14, 45]. The use of Interactive visual labeling (VIAL) systems [9], which are built over active learning algorithms, can improve their performance [8].

Labeling for downstream tasks. There are many approaches for labeling complex, multi-variate data for downstream tasks. Some are more generic frameworks for multi-variate data [6, 20]. Others make custom interfaces for highly specialized data or tasks, such as motion-capture data [7] and image segmentation tasks [1].

In this work, we focus on the visualization of complex, multi-modal HT-specific data where the label generation is for a downstream task.

3 Design Considerations

Our goal is to build an interactive system for HT that lets a domain expert investigate and label possibly suspicious meta-clusters. TrafficVis is the result of 9 months of participatory design [43] in conjunction with domain experts. More specifically, we received weekly feedback from two domain experts: a Senior Research Scientist at *Marinus Analytics* with six years of analyzing escort ads for HT and extensive experience working in government, and an HT survivor and analyst with 20 years of experience helping trafficked minors on the street. We invited both experts to be co-authors on this paper due to their extensive feedback. Both of these domain experts are well-connected with people in law enforcement in the US, both at the local and federal level. Through extensive discussions with these domain experts, we distilled the following design considerations (C1 – C3).

C1. **Big Picture**: *Connect micro-clusters into larger activities.* While traffickers often entirely control the ad content for their victims [21, 27, 31, 50], over time they might make changes to the text or post multiple ad templates at once. Domain experts state that metadata (e.g., phone number, email address) can be useful in connecting micro-clusters into larger organized activities, which we call meta-clusters (formally defined in Section 5).

An important design principle of labeling systems is to allow users to see both high level and low level information [41]. While domain experts are generally interested in the behavior of the meta-cluster as a whole, they also would like the ability to drill down into a particular micro-cluster, particularly if it has different behavior than other micro-clusters in the meta-cluster.

C2. **Multimodality**: *Displaying complex, multimodal data.* With many previous labeling systems, the challenge was label recommendation, rather than visualizing complex, multimodal data. However, in our case, the challenge lies in effectively visualizing spatio-temporal and text data of meta-clusters to domain experts.

Each meta-cluster has many time series. The posting pattern of metadata fields, geographic locations, and micro-clusters can each be represented as a time-series through the lifetime of the meta-cluster. Furthermore, each micro-cluster within the meta-cluster has its own time-series for each of these fields.

Since many domain experts look for suspicious keywords in ads to determine the M.O. [2, 15], thoughtfully showing the text in each meta-cluster is important. In particular, our domain experts often mentioned the need to be able to drill down into particular ad text while still being able to see the overarching text patterns.

Given the importance of promoting rich interactions between the data and domain experts [16], we must enable them to navigate through this data efficiently.

C3. **Usability**: *Making the interface usable for law enforcement:* Since some of our intended users will be law enforcement officers, all plots must be easily understood by non-experts in visualization. How can we convey patterns intuitively, using methods that the average law enforcement officer will understand?

Furthermore, through our domain experts, we've gathered that law enforcement likes to see as much information at once as possible; they do not like applications that require lots of scrolling back and forth to see the results. However, we also must ensure we do not overwhelm the expert [26].

The above design considerations cover the features mentioned to us during our conversations with domain experts.

4 TrafficVis: Backend Algorithm Design

How can we design our backend algorithms to address the design considerations synthesized in Section 3? Here, we will use the same labels, C1-C3, to specifically mention how our algorithms address these considerations.

4.1 InfoShield

We use a very recent HT detection algorithm, InfoShield [28], to create micro-clusters (text-based clusters). InfoShield exploits the insight that similar ads are likely written by the same person. More specifically, InfoShield is comprised of two parts. First, *InfoShield-coarse*, which quickly creates micro-clusters by connecting ads that share common phrases (up to 5-grams) with a high *term frequency inverse document frequency (tf-idf) [24]* score. Then, *InfoShield-fine* uses the Minimum Description Language (MDL) [40] principle to generate a template for each micro-cluster, aligning ads to find similar phrases and highlight the differing ones through insertions, deletions, or substitutions. InfoShield also finds *slots* — portions of the template that differ for most ads. Slots often contain information specific to that ad, such as name, contact time, or available hours. A visual example of this process is shown in Figure 2. InfoShield also ranks micro-clusters using the relative length metric r (compression ratio of the micro-cluster using the calculated template), which we will use in Section 4.3. We chose InfoShield because it is scalable, achieving near-linear performance on the number of ads processed and explainable, which justifies the decision to create the group to investigators.

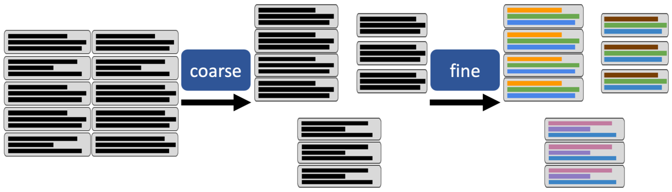


Fig. 2. *Pipeline for InfoShield:* Taking crawled ads as input, InfoShield-coarse groups these ads into micro-clusters, and InfoShield-fine highlights the common phrases in each ad by finding a common template.

4.2 Meta-Clustering: C1 (Big Picture)

Since micro-clusters are constructed only using text features, multiple micro-clusters can actually be part of the same activity. Therefore, we connect micro-clusters (c_i) into *meta-clusters* (M_j) based on extracted metadata – images, emails, phone numbers, and social media accounts. We consider two micro-clusters c_1, c_2 to be part of the same meta-cluster M_j if two ads $a_m \in c_1, a_n \in c_2$ share at least one metadata field. Figure 3 shows an example of how six micro-clusters can be connected into three meta-clusters.

This addresses consideration C1 (Big Picture) since we are connecting micro-clusters into larger organized activities. We consider these metadata to be hard connections because of their nature; it is very unlikely that two unrelated micro-clusters are using the same contact information or the same exact image. If we were using metadata fields where the connections were less certain, running a clustering algorithm on this constructed metadata graph could have been an appropriate next step.

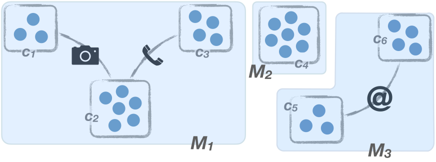


Fig. 3. *From micro-clusters* (c_i) *to meta-clusters* (M_j)*:* By incorporating metadata – images, phone numbers, and social media accounts – we combine 6 micro-clusters into 3 meta-clusters, each of which are part of the same group.

4.3 Ranking: C3 (Usability)

We would like domain experts to both view and label the most suspicious clusters first, so that we can get useful labels while saving law enforcement's time, addressing consideration C3 (Usability). Domain experts consider some suspicious signals to be (a) a large number of ads and micro-clusters in the meta-cluster, and (b) very similar text. Large numbers of ads and micro-clusters are considered suspicious since they hint at the existence of a large organized group, with too many ads and clusters to be an individual escort. Similar text can also be considered a suspicious signal since traffickers often use the same template to advertise many victims. Therefore, we devise a ranking heuristic to prioritize types of meta-clusters with those behaviors. Since we observe the number of ads and number of micro-clusters in meta-clusters to be Pareto distributed, we will scale these values logarithmically. In order to capture text similarity, we consider the relative length metric (that is, compression ratio) r given from InfoShield, which measures the goodness of compression for a particular micro-cluster. r is close to 1 if the compression is bad, and smaller if the compression is good, signifying high similarity among the ads of the micro-cluster. More specifically, for a given meta-cluster, let N be the number of ads, M be the number of micro-clusters, and let $R = \{r_1, r_2, ..., r_M\}$ be the relative length scores for each micro-cluster. We give each meta-cluster a suspiciousness score s by

$$s(N,M,R) = \frac{\log N + \log M}{\frac{1}{M}\sum_{i=1}^{M} r_i}. \tag{1}$$

This metric will prioritize meta-clusters that have a large number of ads and micro-clusters, that also have good compression. Since the compression ratios R are positive and less than 1, micro-clusters with more similar text will boost the score. Finally, we present the meta-clusters in TrafficVis from high to low score. For the few labeled clusters that we have, we do observe that this metric ranks national-level HT rings and *spam* meta-clusters first.

5 TrafficVis: UI Design

Once we construct meta-clusters, how can we visualize them in a way that addresses our design considerations from Section 3? We will describe each part of the interface through a usage scenario. As we introduce the features of TrafficVis, we will annotate which design consideration C1–C3 it addresses. We then present a scenario based on the real experiences of crime analysts, inspired by actual comments given by experts on the presented data during solicited feedback (see Section 6).

5.1 The User Interface

First, we will quickly describe each part of the interface, and further elaborate on how each panel is used in Section 5.2.

The top banner shows basic statistics for the meta-cluster. The *Micro-cluster panel* shows the posting behavior of the top 10 micro-clusters with the highest number of ads throughout the lifetime of the meta-cluster, as seen in Figure 6 (left). This addresses C1 (Big Picture) by allowing users to see the posting behavior of the micro-clusters and the overall meta-cluster simultaneously. By hovering over a particular cell, a tooltip displays the number of ads per day in that micro-cluster. A multi-select dropdown above the *Micro-cluster panel* (Figure 6 (left)) allows users to select a particular subset of micro-clusters, which will update all panels and text in the rest of the interface. This feature further addresses C1 by letting users customize exactly which micro-clusters they can drill down into. By deselecting all micro-clusters, all meta-cluster data will once again be displayed in all panels.

The *Timeline panel* (Figure 4) shows the usage of metadata and the number of micro-clusters with posted ads per day over the lifetime of the meta-cluster. By hovering over any date, the time-series values will be displayed. Since the number of micro-clusters is a feature derived from InfoShield, it is displayed separately. The *Map panel* (Figure 5) also shows the geographic spread of the meta-cluster or selected micro-clusters. A tooltip shows the number of ads posted in each location. These panels help us display complex, spatio-temporal data usefully, addressing C2 (Multimodality).

The *Text panel* (Figure 6) allows domain experts to scroll through the text templates generated by InfoShield, as shown in Figure 6 (top), which give a general sense of the phrasing for each micro-cluster. If any micro-clusters are selected, the *Text panel* will show the individual ads for those particular micro-clusters, as shown in Figure 6 (bottom), highlighting any deviations from the template as insertions, substitutions, or slots, as designated by InfoShield. This panel helps us display complex text data usefully and drill down into individual micro-clusters when needed (C1, C2).

The *Labeling panel* (Figure 1.5) lets the domain expert quickly label the meta-cluster on a scale of 1 (very unlikely) to 5 (very likely) for each possible M.O. We use sliders and a discrete scale, rather than a continuous input probability, for ease of use, addressing C3 (Multimodality). Upon clicking the 'Next meta-cluster' button, these labels are saved to a CSV file and a new meta-cluster is displayed.

5.2 Usage Scenario: Analyst finding a massage parlor cluster with suspected HT

We present a usage scenario to illustrate how each panel can work together to help an analyst (e.g., law enforcement agent, HT domain expert) use TrafficVis to investigate a meta-cluster. This scenario is based on expert feedback solicited on the meta-cluster depicted in Figures 1, 4, 5, and 6.

First, the analyst sees high-level statistics on the top banner, observing that for this 200 ad cluster, there are a lot of images posted, which often correlates with organized behavior. She then moves to the *Micro-cluster panel* (Figure 6 left) to inspect the individual micro-clusters. The analyst may choose to further investigate the consistent volume of ads in micro-cluster c_1 during the last few months of the meta-cluster using the multi-select dropdown just above the *Micro-cluster panel*. When she does, the entire interface will populate with that micro-cluster's geographic, temporal, and text data.

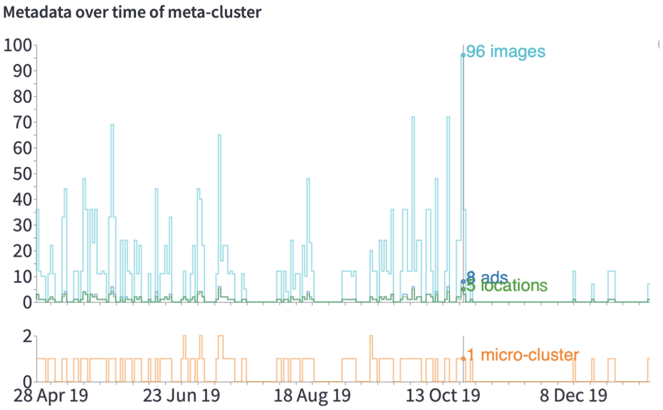


Fig. 4. **Irregular spikes in *Timeline panel***, indicating to experts that this is not script-generated posting behavior, but rather human-generated, ruling out *spam* and *scam* labels.

Next, the analyst can look at metadata usage and the number of micro-clusters per day in the *Timeline panel* (Figure 4). She may notice the somewhat inconsistent posting over many months, with some "hot spots". This does not look like script-generated posting behavior, which makes the label *spam* less likely. She may notice that there are few unique locations per day, which also supports that the ads are not scripted. Using the *Map panel* as shown in Figure 5, she can investigate which locations are most popular, noticing that there is some regional spread focused on bigger cities in the Midwest and East Coast. This could be indicative of *trafficking* or *benign* behavior, with an individual or group circling between cities likely to have many customers, which rules out *spam* and *scam*.

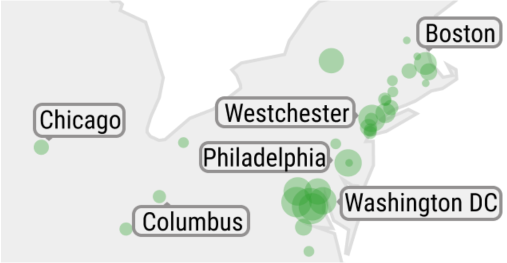


Fig. 5. **Meta-cluster focuses on big cities**: ads are focused on bigger cities in the Midwest and East Coast. Size represents the number of ads posted in that location. This could be indicative of *trafficking* or *benign* workers circling between cities with many customers.

Next, she inspects the text of each ad in the scrollable *Text panel* (Figure 6). If a particular meta-cluster is not selected, then only the InfoShield templates for each micro-cluster will be shown. This way, she can compare the differences between the wording in these micro-clusters. By selecting a micro-cluster, then the *Text panel* will change, showing the template and the individual ads for that particular micro-cluster, and highlighting where the individual ad text differs from the template. As found by InfoShield, blue represents insertions, deletions, and substitutions, and red represents parts of the ad that differ from most other ads in the micro-cluster. Looking at c_2, the analyst might notice the following interesting text features (T1-T4), as annotated in Figure 6.

T1. **Social media presence:** sign of a legitimate person (*trafficking* or *benign*). Handles blurred for privacy.

T2. **Preferences:** asks for cleanliness and no unsolicited photos: likely *benign* sex work, but *trafficking* still possible.

T3. **Different dates & locations:** possible sign of traveling *trafficking* or *benign* sex worker.

T4. **High prices:** usually a sign of a popular *benign* worker.

Finally, after inspecting the information in each panel, the analyst uses the *Labeling panel* (Figure 1.5) to confidently label the meta-cluster. Given the regional geographic spread and that many signals point to the ad not being script generated, the analyst labels this meta-cluster as likely *benign* (an at-will sex worker), with a small chance of *trafficking* based on the suspicious keyword.

5.3 Iterative Design Process

TRAFFICVIS was developed through 9 months of participatory design. We started our design by discussing the interest of visualization for fighting HT with our domain experts, getting their sense of what the basic needs of crime analysts and law enforcement would be. Then, over the span of a couple of weeks, we iterated over a few possible sketches of TRAFFICVIS. Each week, domain experts would give us feedback that would change our final design. As the sketches were implemented on real data, we iterated over many possible encodings of the data with domain experts, who constantly provided us with the perspective of both crime analysts and law enforcement officers. We give some examples of iterations for a few panels below.

5.3.1 *Micro-cluster panel*

We considered network-based representations of the data; nodes would be micro-clusters or individual ads, and edges would be metadata fields or keywords. However, we decided against it due to the hairball effect – most ads are connected with similar keywords and metadata, resulting in many clique-like structures. Since the relationships between micro-clusters were not useful to show, we focused on the timelines between each micro-cluster, as is shown in *Micro-cluster panel*. This helps us address C1 (Big Picture), since users are able to quickly decide if they want to drill down into lower-level information about each micro-cluster.

5.3.2 *Text panel*

Here, we display the results of InfoShield, which detects five types of regions; *constant strings* which are the same in most ads, *insertions, deletions, and substitutions*, and finally *slots*, or places where the text differs in most ads. Originally, we considered displaying output the same way as described in InfoShield, where constant strings were highlighted in yellow, insertions in green, deletions in grey, substitutions in blue, and slots in red. However, this representation became very visually crowded, and the average law enforcement officer does not care about the differences between slots and insertions. Therefore, we changed the representation to not highlight constant strings, make insertions/substitutions/deletions light blue, and slots red. This helps with C3 (Usability), since we are making the design less complex for domain experts that do not need this specific information.

Also, we were originally displaying all of the text for all micro-clusters in this scrollable panel, which ended up being overwhelming and cumbersome for domain experts, since they had to scroll down very far to get to some ads. Instead, we decided to only show the text templates from InfoShield if no micro-cluster was selected, and let the user decide which actual ads they wanted to see by using the multi-selector. This way, we make it easy to drill down into actual ads, addressing C1 (Big Picture).

5.3.3 Drilling down into specific micro-clusters

Domain experts specifically asked for this feature in order to drill down into specific mirco-clusters if needed. At first, we added a selector to *Micro-cluster panel* itself, which enabled users to click on a particular row to select a micro-cluster. However, this ended up being unintuitive, since domain experts would not realize that anything would happen if they did click on each row, so this feature was not being used. Therefore, we created an explicit dropdown with explanatory text. Also, in this way, it was easy to implement mutli-selection without asking the domain expert to remember keyboard shortcuts. Since some law enforcement officers may not be as familiar with common keyboard shortcuts, this makes it easier for them to use TRAFFICVIS (C3).

6 EXPERT FEEDBACK

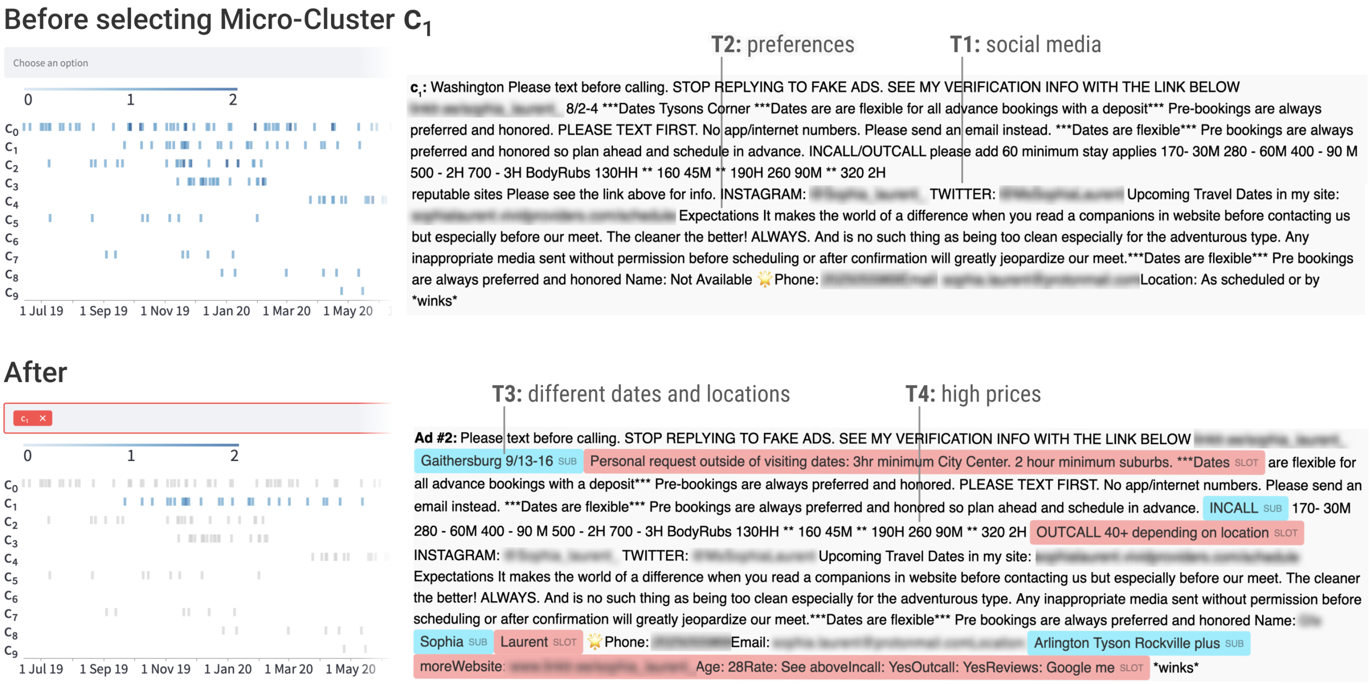
We solicited expert feedback from domain experts to answer the following questions about the efficacy of TRAFFICVIS for inspecting and labeling suspicious meta-clusters. Q1 – Q3 directly correspond to our Design Considerations C1 – C3 introduced in Section 3.

Q1. Evaluating C1 (Big Picture): Is the distinction between meta-clusters and micro-clusters useful to experts? How do they interact with the clustering results?

Q2. Evaluating C2 (Multimodality): How do experts interact with metadata plots? How do these plots influence their labeling decisions?

Q3. Evaluating C3 (Usability): Do the solicited experts, which have varying backgrounds, all find it easy to use? How do they expect other types of experts would react to the design (i.e. law enforcement)?

Q4. Which features of TRAFFICVIS are most important to experts? Are there any *insights about the labeling process* that we gain from seeing how experts look at the data?



c_1: Washington Please text before calling. STOP REPLYING TO FAKE ADS. SEE MY VERIFICATION INFO WITH THE LINK BELOW [redacted] 8/2-4 ***Dates Tysons Corner ***Dates are are flexible for all advance bookings with a deposit*** Pre-bookings are always preferred and honored. PLEASE TEXT FIRST. No app/internet numbers. Please send an email instead. ***Dates are flexible*** Pre bookings are always preferred and honored so plan ahead and schedule in advance. INCALL/OUTCALL please add 60 minimum stay applies 170- 30M 280 - 60M 400 - 90 M 500 - 2H 700 - 3H BodyRubs 130HH ** 160 45M ** 190H 260 90M ** 320 2H
reputable sites Please see the link above for info. INSTAGRAM: [redacted] TWITTER: [redacted] Upcoming Travel Dates in my site: [redacted] Expectations It makes the world of a difference when you read a companions in website before contacting us but especially before our meet. The cleaner the better! ALWAYS. And is no such thing as being too clean especially for the adventurous type. Any inappropriate media sent without permission before scheduling or after confirmation will greatly jeopardize our meet.***Dates are flexible*** Pre bookings are always preferred and honored Name: Not Available 🌟Phone: [redacted]Location: As scheduled or by *winks*

Ad #2: Please text before calling. STOP REPLYING TO FAKE ADS. SEE MY VERIFICATION INFO WITH THE LINK BELOW [redacted] Gaithersburg 9/13-16 SUB Personal request outside of visiting dates: 3hr minimum City Center. 2 hour minimum suburbs. ***Dates SLOT are flexible for all advance bookings with a deposit*** Pre-bookings are always preferred and honored. PLEASE TEXT FIRST. No app/internet numbers. Please send an email instead. ***Dates are flexible*** Pre bookings are always preferred and honored so plan ahead and schedule in advance. INCALL SUB 170- 30M 280 - 60M 400 - 90 M 500 - 2H 700 - 3H BodyRubs 130HH ** 160 45M ** 190H 260 90M ** 320 2H OUTCALL 40+ depending on location SLOT INSTAGRAM: [redacted] TWITTER: [redacted] Upcoming Travel Dates in my site: [redacted] Expectations It makes the world of a difference when you read a companions in website before contacting us but especially before our meet. The cleaner the better! ALWAYS. And is no such thing as being too clean especially for the adventurous type. Any inappropriate media sent without permission before scheduling or after confirmation will greatly jeopardize our meet.***Dates are flexible*** Pre bookings are always preferred and honored Name: [redacted] Sophia SUB Laurent SLOT 🌟Phone: [redacted]Email: [redacted] Arlington Tyson Rockville plus SUB moreWebsite: [redacted] Age: 28Rate: See aboveIncall: YesOutcall: YesReviews: Google me SLOT *winks*

Fig. 6. **Drilling down into specific micro-clusters**: annotations correspond to *text features* T1-T4 from Section 5.2. **Top:** *Micro-cluster panel* shows the posting activity for all micro-clusters. *Text panel* shows the template text for micro-cluster c_1. **Bottom:** upon selecting micro-cluster c_1, *Micro-cluster panel* updates to highlight c_1 and *Text panel* shows the individual ads with differences highlighted. As found by InfoShield, blue highlights represent substituted phrases, and red highlights represent parts of the ad that differ from most other ads in the micro-cluster (known as *slots*). Some sensitive text is blurred.

Q5. How quickly can experts label meta-clusters? Do they believe it will significantly *speed up the labeling process*?

6.1 Intuition Behind Setup

There are few domain experts in HT that analyze escort ads. This not only makes it challenging to solicit them for feedback, but also tasks us with making our study as efficient as possible as to make the best use of their very limited time. We did not want each expert to take more than approximately an hour of their time giving feedback. Since the current solution for domain experts is manual labeling, we decided there is little point in A/B testing. Any clustering and visualization would provide speedup vs. manual labeling. Instead, we focus on asking experts to label more meta-clusters and soliciting feedback on the tool as a whole.

6.2 Solicited experts

We asked domain experts with various experience in HT to participate. We recruited four experts; For brevity, we will use E1, E2...E4 to refer to Expert 1, Expert 2, ...Expert 4. E1 and E2 are from *Marinus Analytics*, a Pittsburgh-based startup focused on fighting HT and the providers of our data. E3 is an HT survivor that now helps rescue trafficked minors on the street. E4 is a criminology master's student studying escort advertisements and HT. E2 and E3 were invited to be co-authors on this paper due to their extensive feedback throughout the development of TRAFFICVIS.

The average time our experts were involved in studying HT varied greatly, from 1 to 20 years. On average, experts rated themselves as a 4.3 ± 1.15 out of 5 on their expertise in labeling escort ads, and a 3.6 ± 0.58 out of 5 on their experience with AI and clustering algorithms.

Some of our domain experts had extensive experience with looking for cases themselves. E2 and E3 discussed using keyword searches and various statistical techniques to investigate clusters. In their experience with law enforcement, local officers often start with a tip and try to build a case manually using street knowledge. Experts say some officers may look for online ads to try and glean some information, but it's difficult to find in a large set of unorganized ads on a webpage.

6.3 Dataset used

We used a random sample of escort ads given to us by *Marinus Analytics*. These ads were crawled from multiple common escort websites which are suspected to contain organized activity. Then, we ran InfoShield on these ads, followed by our meta-clustering algorithm. We then manually picked 10 meta-clusters that differed in their spatio-temporal distributions, number of posted ads, and text templates (i.e. length, use of emojis, presence of suspected *trafficking* keywords), to increase the likelihood of differing labels. We chose 10 in an effort to limit the time taken for each interview to no more than one hour.

6.4 Procedure

Our protocol was approved by the IRB. Each expert signed a consent form before the interview was conducted. All experts were interviewed separately over Zoom. All experts' movements on the interface and their audio were recorded. Experts got access to the interface by the interviewer sharing their screen and letting the expert interact with it using Zoom's remote control feature. Each interview started with a 5-minute introduction, outlining the structure of the interview. Then, each expert was asked the questions about their background in HT, analyzing escort ads, and whether they have any insights to law enforcement. The exact wording of these questions can be found in the supplemental material.

We then gave a 5 minute tutorial on TRAFFICVIS, making sure the expert had common definitions for all M.O.s. Then, we let the expert explore the interface using Zoom remote control and label the 10 clusters in our dataset. The labeling options were 1: Very unlikely, 2: Unlikely, 3: Unsure, 4: Likely, and 5: Very likely. Experts were encouraged to think aloud as questions and comments arose, and to verbally explain the clues that led them to their final decision. If they had not previously explained their thought process, upon finalizing their labels for a meta-cluster, the interviewer would ask them to quickly explain why. We recorded the elapsed time to label each meta-cluster as the moment the interface loaded a new meta-cluster to the moment they clicked the 'next meta-cluster' button.

Once experts finished labeling all clusters, we would end the session with a few exit questions asking for feedback about TRAFFICVIS, which can be found in the supplemental material. Finally, the expert was asked to complete a quick questionnaire offline, which can also be found in the supplemental material.

6.5 Results and Design Lessons

Experts had overwhelmingly positive feedback on TRAFFICVIS. They predominantly looked at the text to identify the behavior of clusters, but used the geographic spread and timelines to supplement their thinking. Often, specific keywords would jump out at them. Based on their feedback, we distilled some central design lessons. We show the number of experts that commented on each lesson, without being prompted, in Figure 7. Next, we elaborate on the design lessons we learned (L1 – L6) and how they answer our questions Q1 – Q6.

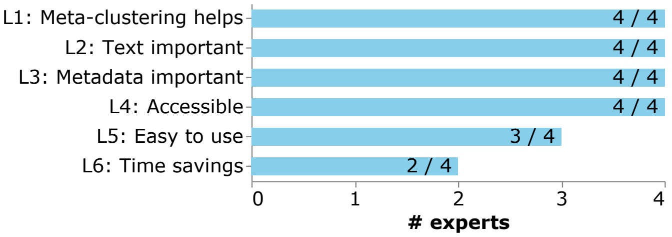


Fig. 7. **Design Lessons from Expert Feedback:** the number of experts that explicitly commented on each design lesson, without being prompted.

L1: Meta-clustering helps – (Q1) All experts enjoyed that the ads were clustered together, rather than looking at them individually as is typically done in the current approach, manual labeling. The distinction between micro-clusters and meta-clusters was also appreciated by the experts. E1 mentions that they

> "*liked the ability to look at the meta-cluster and then select some of the micro-clusters and see some of the ads within those...[to] be able to drill down*".

E3 spent much of their time drilling down into particular micro-clusters to see how they varied, saying

> "*being able to look at the individual micro-clusters really helps.*"

E4 also mentioned the distinction between meta and micro-clustering as useful, saying that they liked

> "*the fact that you could go within certain ads within the micro-clusters*"

L2: Text is important – (Q2, Q4) Text was the most defining feature. All experts really enjoyed the text clustering, spending the majority of their time interacting with TRAFFICVIS on the *Text panel*. E4 says they

> "*liked the ad text...I used that a lot in how I was labeling.*"

For each meta-cluster, every expert cited text as part of their reasoning for selecting their labels. For most meta-clusters, all experts looked at the text the majority of the time. In particular, the presence or absence of certain keywords were telling. When looking at the first meta-cluster, E2 mentions

> "*I'm looking for keywords I usually see when I look at this stuff*".

We distilled the most commonly cited text indicators of *trafficking*.

- Mention of exotic ethnicities (E1, E3)
- Scarcity: girls are constantly changing, new in town, or leaving soon (E1, E3)
- Multiple girls advertised (E1, E2)
- Particular keywords, redacted by request of the participating domain experts (E1 – E4)
- Offering high-risk services (E1 – E4)

Experts also commented that short ads or differing fonts in ads are possible indicators of *spam* or *scam* (E1, E2, E4).

L3: Metadata is important – (Q2, Q4) it provides useful insights. Experts commonly referenced various metadata properties as influencing their labeling decisions. All experts interacted with all panels during their labeling. E1, E2, E3 explicitly mentioned looking at the number or similarity of phone numbers and E1-E4 all mentioned the geographic spread when explaining their rationale for labeling particular meta-clusters. E3 referenced the temporal distribution multiple times, saying

> "*since you have multiple ads a day, this is likely not an individual escort...[they] post a ton of websites on one day, but spread over a year*" as part of their reasoning."

E4 particularly liked seeing the geographic spread:

> "*the geographical spread is very good too, it's a very good indicator especially when you're labeling spam and scam, and also trafficking...*"

Experts commented that seeing an explicit location offered in the ad text is an indicator against spam or scam (E2, E3, E4).

L4: Accessible – (Q3) TRAFFICVIS is useful for many anti-HT stakeholders. All experts commented that TRAFFICVIS could be used by domain experts and law enforcement. E1 believes that TRAFFICVIS is

> "*really powerful for finding large organized crime groups.*"

E1 suggested it would likely be

> "*more relevant to larger national law enforcement groups than local [law enforcement], but [they] think it would be helpful in building cases or showing relationships.*"

E2 commented that they

> "*could really see this speeding up scanning ads, especially if you've got one cluster of ads and you see another cluster in the same geographical area, even over the same time period.*"

E3 mentioned that a huge benefit of TRAFFICVIS was the curation of meta-cluster labels, stating that TRAFFICVIS is

> "*useful for labeling, for supervised training which is currently very difficult...and also for verifying whether the underlying algorithms are correct.*"

E3 was particularly excited about the possibility for law enforcement and law attorneys to use TRAFFICVIS, stating that

> "*it would help investigators retrace their steps from jury or prosecution during testimony...it would really add to explainability and justification for why that individual is being indicted...[or] targeted.*"

E2, E3, and E4 all explicitly mentioned that they'd like to label more clusters with TRAFFICVIS.

L5: Easy to use – (Q3) experts thought TRAFFICVIS was well-designed. Broadly, experts liked the interface, saying they were "very impressed by the tool" (E2, E4) and that it's "easy to use" (E3, E4). E2 and E3 particularly enjoyed the one page layout, mentioning that "you didn't have to jump to another page to record your responses" and "the layout is nice, don't change it".

L6: Time savings – (Q5) TRAFFICVIS makes labeling possible. E3 commented on the possible time savings as compared to Marinus' escort ad exploration software:

> *"it's quick...even with TrafficJam it would take 20-30 minutes per cluster to try and figure out what is going on. And then you wouldn't even be able to label the ads. This is a huge advantage over the way things are currently done. For law enforcement officers they have no way...they have to do everything manually, there's no way."*

E2 also mentioned the time savings, saying

> *"I could really see this speeding up scanning ads, especially if you've got one cluster of ads and you see another cluster in the same geographical area, even over the same time period."*

Experts consistently need about 2-3 minutes to provide labels with TRAFFICVIS, with an average of 2 minutes and 36 seconds. The distributions of the time taken to label, per expert and per meta-cluster, are shown in Figure 8, in comparison to E3's estimation of 20-30 minutes.

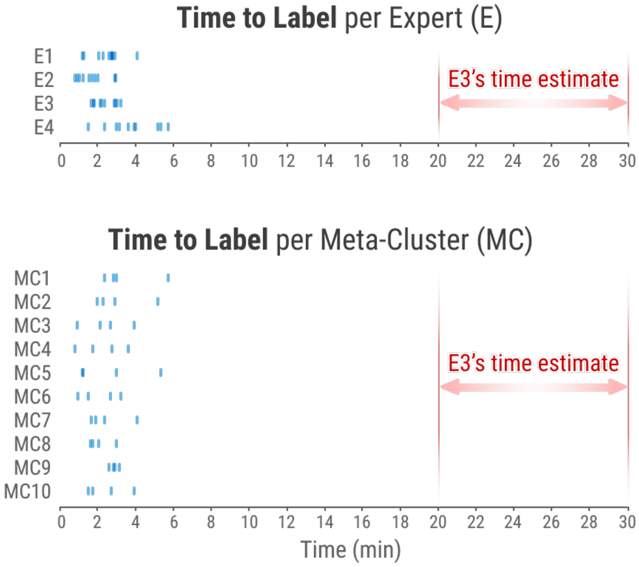


Fig. 8. **TRAFFICVIS is fast:** Experts consistently need about 2-4 minutes to provide labels, while E3 estimates any other method would take at least 20-30 minutes. (top) labeling times by expert, (bottom) shows labeling times by meta-cluster.

This confirms a central motivating point of TRAFFICVIS: the current solution, manual labeling, is so time-intensive, that it is rarely ever done. With TRAFFICVIS, it's *feasible* to solicit labels from domain experts.

6.6 Distribution of labels

The final labels, averaged among all experts, are shown in Figure 9 for each meta-cluster. Circles represent the predominant label for each meta-cluster, while tick marks represent all other labels. For example, in MC1, our experts gave an average score of 1, 1.25, 3, 2.5 and 4.25 to *spam*, *scam*, *trafficking*, *benign* and *massage parlor* labels respectively. This strongly indicates that MC1's most likely label is *massage parlor*. Looking at the highest-valued label for each meta-cluster (circles), we see that we ended up with 6 *benign*, 3 *trafficking*, and 1 *massage parlor*.

6.6.1 Post-interview Questionnaire

The results of the post-interview questionnaire can be seen in Figure 10. We note that all users had a positive experience with TRAFFICVIS and see it implemented in practice.

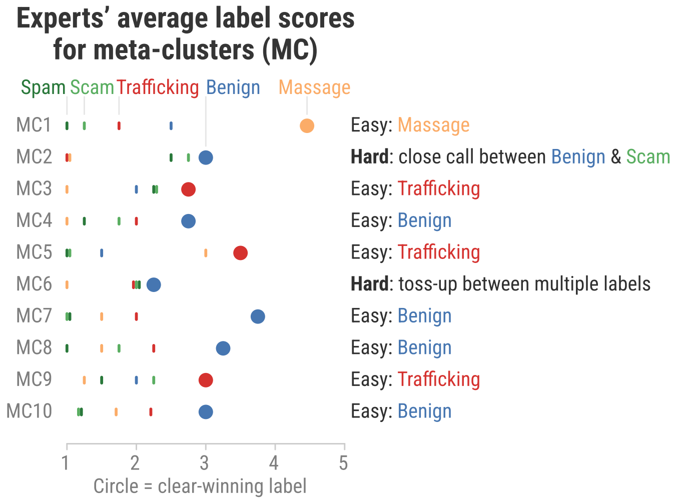


Fig. 9. **Final labels**: averaged scores among all experts, for each meta-cluster. Circles represent clear winning labels. Experts usually agreed on one label, except for a few meta-clusters that are close calls (2, 6). For both these meta-clusters, at least one expert called it difficult to label based on the given information.

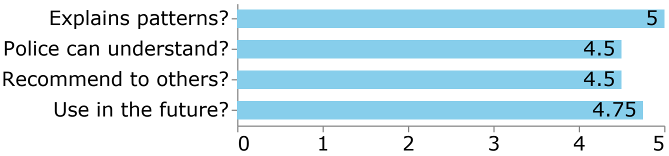


Fig. 10. **Experts loved TRAFFICVIS:** results on a scale of 1 (strongly disagree) to 5 (strongly agree). Full questions in supplemental material.

7 LIMITATIONS AND FUTURE WORK

7.1 Improvements to Algorithms and UI Design

Experts seemed confident in the results of our algorithms. However, we would like to integrate more features into TRAFFICVIS, such as an automated analysis of the spatial trajectories of meta-clusters over time. If a particular meta-cluster shows ads moving across the US over time or circling back to the same few locations again and again, this could be indicative of a traveling HT ring. Experts could use these patterns to better inform their labeling.

While we don't currently have access to sensitive image data, only the hash codes for those images, experts often look at image data to inform their labels. In particular, E3 noted that including image data would have made it easier to label some meta-clusters (i.e. MC2, MC6). If we get access to image data in the future, TRAFFICVIS could analyze it to help experts in various ways. For example, we could estimate the number of distinct people advertised in a particular meta-cluster, where a large number of possible victims would be indicative of an organized HT ring.

In terms of the UI design, we got two minor pieces of feedback. E2 asked for larger font sizes and E3 was interested in adding an additional label: "possible", which would fall between "3: Unsure" and "4: Likely". Any substantial improvements to UI design would arise if we implemented some of the algorithmic improvements mentioned above.

In the future, we would also like to incorporate more than just escort ads in our analysis. Many ads have connections to social media websites, such as Twitter, Instagram, Facebook, and OnlyFans accounts. We would like to collect some of this data and incorporate it into our algorithms and visualizations. In particular, Instagram and Facebook are the most common platforms for soliciting escort services [36], but their terms of service do not allow for crawling data. It's also against

OnlyFans terms of service to crawl data, leaving only Twitter as a possible source. Since we see few Twitter handles mentioned in escort ads compared to the other aforementioned platforms, we believe the benefit of incorporating Twitter data would be marginal at best.

7.2 Soliciting Additional Feedback from Law Enforcement

We are interested in feedback from law enforcement officers. Unfortunately, they are busy and not always willing to evaluate tools. Since our broader goal is to help them more quickly find actionable HT cases, we are connecting with *Marinus Analytics*, who has direct ties to law enforcement. We expect this to increase the visibility of TRAFFICVIS, and more broadly, our work in anti-HT. We hope that in the future, this can be used not only to label, but to empower law enforcement to quickly investigate any possibly suspicious clusters of escort ads.

7.3 Societal Impact and Practical Use

There has been much backlash in the media as well as academia about the use of black box technologies in law enforcement [4,35]. Frequently, law enforcement efforts are based on predictions of illegal activity derived from AI or other algorithms that do not provide explainability and that prosecuting attorneys do not understand. Since TRAFFICVIS is designed to detect and visualize organized crime groups, it has the potential to be an ideal tool to explain how one arrived at the decision that a case was a part of an organized crime group.

According to E3, TRAFFICVIS could be justifiably used in court because no black box algorithms were utilized. The visual presentation of individual ads in the *Text panel* shows exactly how the ads are connected. For example, in Figure 1, we can see that the majority of the ad content is the same, excepting some details such as dates and locations.

One of the largest concerns we have with building algorithms and systems for fighting HT is to make sure that we are not stepping on the liberties of at-will sex workers, who also post escort ads on these websites. While large clusters of text similarity generally signal organized crime groups and not individual workers, we are conscious that they may appear in a meta-cluster. Since TRAFFICVIS does not outwardly classify any meta-clusters as HT cases, only highlighting some possibly suspicious ones, we put the onus on domain experts to make the final decision.

However, even without TRAFFICVIS, a law enforcement officer could look at any of these ads online, set up a fake appointment with a real escort worker, and arrest them for prostitution at any time. We have to be very careful about which officers will get access to this software and data, and we are working with *Marinus Analytics* to ensure that anyone with direct access to a running instance of TRAFFICVIS would be highly vetted. Furthermore, we've seen an encouraging trend towards law enforcement taking a victim-centered approach to HT; many cities have been decriminalizing prostitution in the past year [5,34]. By vetting the law enforcement users of TRAFFICVIS to only officials that are clearly invested a victim-centric approach, we ensure that TRAFFICVIS does not contribute to stepping on the liberties of at-will sex workers. We also have ensured that we have stakeholders in multiple affected populations: not just the perspective of *Marinus Analytics* and law enforcement, but also of HT survivors. One of the domain experts, who is a coauthor on this paper, is a survivor of HT who now helps trafficked minors on the street, and their perspective strongly informed the design of TRAFFICVIS.

7.4 Using our labels for downstream tasks.

The labeling design of TRAFFICVIS was intentionally chosen to be flexible for the expert, allowing them to rate on a scale of 1–5 for each label. However, this causes some difficulties for us to post-process these labels before they are used in downstream tasks, particularly because the labels are not disjoint. Downstream classification of meta-clusters will be difficult since the same meta-cluster could be labeled in 25 different ways, and not all differences between labels are insightful – the difference between a '4: Likely' and '5: Very Likely' may not be very meaningful. Furthermore, a meta-cluster could simultaneously be *benign* and *massage parlor*, or *trafficking* and *massage parlor*. To handle this, we can do a few different things: (a) threshold the label scores, i.e. an average score of 3.5 or higher indicates the meta-cluster falls under that label, else it does not, (b) choose to not predict certain labels that overlap with others, i.e. *massage parlor*, or (c) treat our downstream task as prediction rather that classification. We are in active discussion with our domain experts to decide what the best next steps will be.

7.5 Reproducibility and Application to Other Domains

Unfortunately, we cannot make the data publicly available to protect the safety of potential victims. However, even with public data, our study could only be reproduced by somebody able to solicit HT experts. Within these parameters, we have done what we can to make TRAFFICVIS reproducible; the code for InfoShield and TRAFFICVIS are open-sourced with synthetic data.

TRAFFICVIS has specifically been designed for labeling suspicious meta-clusters of escort ads for HT and other organized activity. However, TRAFFICVIS could be applied as a cluster labeling solution for other domains. For example, coordinated disinformation campaigns on social media have become a pervasive issue in the last few years [10,47], causing many tech companies to implement algorithms to find and flag suspicious users online [17]. One could use TRAFFICVIS's pipeline to quickly label clusters of similar social media posts, using relevant metadata such as images and usernames. In fact, the same clustering algorithm could possibly be used; InfoShield was also found to be successful in a social media context – finding organized, bot-like behavior in Twitter data [28].

8 CONCLUSIONS

Facilitating the retrieval of high-quality labels for complex, multimodal data can be a challenging task. TRAFFICVIS is a system designed to visualize this type of data for the HT problem, making the following major contributions:

1. **High-impact,** being accessible to a variety of anti-HT stakeholders, including criminologists, domain experts, and law enforcement (see Section 6);

2. **Label generation**, finally providing a way to generate high-quality cluster labels, which will be used for further algorithm development;

3. **Time-saving**, granting a huge speedup over manual labeling, according to feedback from domain experts.

TRAFFICVIS shows that even with such complex data, we can still design an interface that lets domain experts quickly see big patterns while simultaneously allowing them to drill down into specific entries when needed.

Through the process of soliciting expert feedback, we naturally curated a dataset labeled by TRAFFICVIS that will enable further algorithm development towards M.O. detection, allowing law enforcement to quickly find meta-clusters of ads that actually represent real HT cases. We plan to have interested experts continue to label more meta-clusters using TRAFFICVIS. *Marinus Analytics* has expressed interest in incorporating TRAFFICVIS in their pipeline, which would allow us to continue getting more labeled clusters. This process will enable researchers to continually develop and evaluate M.O. detection algorithms as we see emerging trends in escort ads over the years to come. Using the labels generated by TRAFFICVIS, we can now start to develop and evaluate novel M.O. detection methods that can further help law enforcement; by removing *spam* clusters, we can increase the rate at which they will find and pursue actual HT cases.

Reproducibility: The code and synthetic data is open-sourced at `https://github.com/catvajiac/TrafficVis`.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE1745016 and DGE2140739.

References

[1] D. Acuna, H. Ling, A. Kar, and S. Fidler. Efficient interactive annotation of segmentation datasets with polygon-rnn++. In *CVPR*, pp. 859–868. Computer Vision Foundation / IEEE Computer Society, 2018. 3

[2] H. Alvari, P. Shakarian, and J. E. K. Snyder. A non-parametric learning approach to identify online human trafficking. In *ISI*, pp. 133–138. IEEE, 2016. 3

[3] H. Alvari, P. Shakarian, and J. K. Snyder. Semi-supervised learning for detecting human trafficking. *Security Informatics*, 6(1):1, 2017. 2

[4] Y. Bathaee. The artificial intelligence black box and the failure of intent and causation. *Harvard Journal of Law & Technology*, 31:889, 2018. 9

[5] J. Battaglia. Baltimore will no longer prosecute drug possession, prostitution and other low-level offenses. `https://www.cnn.com/2021/03/27/us/baltimore-prosecute-prostitution-drug-possession/index.html`, 2021. 9

[6] D. Beil and A. Theissler. Cluster-clean-label: an interactive machine learning approach for labeling high-dimensional data. In *VINCI*, pp. 5:1–5:8. ACM, 2020. 3

[7] J. Bernard, E. Dobermann, A. Vögele, B. Krüger, J. Kohlhammer, and D. W. Fellner. Visual-interactive semi-supervised labeling of human motion capture data. In *Visualization and Data Analysis*, 2017. 3

[8] J. Bernard, M. Hutter, M. Zeppelzauer, D. W. Fellner, and M. Sedlmair. Comparing visual-interactive labeling with active learning: An experimental study. *IEEE Trans. Vis. Comput. Graph.*, 24(1):298–308, 2018. 2, 3

[9] J. Bernard, M. Zeppelzauer, M. Sedlmair, and W. Aigner. A unified process for visual-interactive labeling. In *EuroVA EuroVis*, 2017. 2, 3

[10] D. A. Broniatowski, A. Jamison, S. Qi, L. Alkulaib, T. Chen, A. Benton, S. C. Quinn, and M. Dredze. Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American Journal of Public Health*, 108:1378–1384, 2018. 9

[11] M. D. Buhrmester, T. N. Kwang, and S. D. Gosling. Amazon's mechanical turk. *Perspectives on Psychological Science*, 6:3 – 5, 2011. 2

[12] M. Chegini, J. Bernard, P. Berger, A. Sourin, K. Andrews, and T. Schreck. Interactive labelling of a multivariate dataset for supervised machine learning using linked visualisations, clustering, and active learning. *Vis. Informatics*, 3:9–17, 2019. 3

[13] M. Chen and A. Golan. What may visualization processes optimize? *IEEE Trans. Vis. Comput. Graph.*, 22(12):2619–2632, 2016. 2

[14] M. Desmond, M. J. Muller, Z. Ashktorab, C. Dugan, E. Duesterwald, K. Brimijoin, C. Finegan-Dollak, M. Brachman, A. Sharma, N. N. Joshi, and Q. Pan. Increasing the speed and accuracy of data labeling through an AI assisted interface. In *IUI*, pp. 392–401. ACM, 2021. 3

[15] A. Dubrawski, K. Miller, M. Barnes, B. Boecking, and E. Kennedy. Leveraging publicly available data to discern patterns of human-trafficking activity. *Journal of Human Trafficking*, 1(1):65–85, 2015. 3

[16] J. J. Dudley and P. O. Kristensson. A review of user interface design for interactive machine learning. 8(2), June 2018. doi: 10.1145/3185517 3

[17] C. Duffy. Youtube is cracking down on anti-vaccine misinformation. `https://www.cnn.com/2021/09/29/tech/youtube-vaccine-misinformation/index.html`, 2021. 9

[18] S. S. Esfahani, M. J. Cafarella, M. B. Pouyan, G. J. DeAngelo, E. Eneva, and A. E. Fano. Context-specific language modeling for human trafficking detection from online advertisements. In *ACL*, 2019. 2

[19] K. Fort, G. Adda, and K. B. Cohen. Last words: Amazon mechanical turk: Gold mine or coal mine? *Computational Linguistics*, 37:413–420, 2011. 2

[20] A. P. Hinterreiter, P. Ruch, H. Stitz, M. Ennemoser, J. Bernard, H. Strobelt, and M. Streit. Confusionflow: A model-agnostic visualization for temporal analysis of classifier confusion. *IEEE Trans. Vis. Comput. Graph.*, 28(2):1222–1236, 2022. 3

[21] K. Hundman, T. Gowda, M. Kejriwal, and B. Boecking. Always lurking: Understanding and mitigating bias in online human trafficking detection. In *AIES*, pp. 137–143. ACM, 2018. 3

[22] Ilo global estimate of forced labour. `https://www.ilo.org/wcmsp5/groups/public/@dgreports/@dcomm/documents/publication/wcms_575479.pdf`, 2017. 2

[23] P. G. Ipeirotis, F. J. Provost, and J. Wang. Quality management on amazon mechanical turk. In *HCOMP '10*, 2010. 2

[24] K. S. Jones. A statistical interpretation of term specificity and its application in retrieval. *J. Documentation*, 60(5):493–502, 2004. 3

[25] M. Kejriwal, J. Ding, R. Shao, A. Kumar, and P. A. Szekely. Flagit: A system for minimally supervised human trafficking indicator mining, 2017. 2

[26] T. Kulesza, M. M. Burnett, W. Wong, and S. Stumpf. Principles of explanatory debugging to personalize interactive machine learning. In *IUI*, pp. 126–137. ACM, 2015. 3

[27] A. Kulshrestha. *Detection of Organized Activity in Online Escort Advertisements*. PhD thesis, 2021. 3

[28] M. Lee, C. Vajiac, A. Kulshrestha, S. Levy, N. Park, C. Jones, R. Rabbany, and C. Faloutsos. INFOSHIELD: generalizable information-theoretic human-trafficking detection. In *ICDE*, pp. 1116–1127. IEEE, 2021. 2, 3, 9

[29] L. Li, O. Simek, A. Lai, M. P. Daggett, C. K. Dagli, and C. Jones. Detection and characterization of human trafficking networks using unsupervised scalable text template matching. In *IEEE BigData*, pp. 3111–3120. IEEE, 2018. 2

[30] Marinus Analytics. `www.marinusanalytics.com`. 2

[31] S. D. MINOR. A report on the use of technology to recruit, groom and sell domestic minor sex trafficking victims. 2015. 2, 3

[32] C. Nagpal, K. Miller, B. Boecking, and A. Dubrawski. An entity resolution approach to isolate instances of human trafficking online. In *NUT@EMNLP*, pp. 77–84. Association for Computational Linguistics, 2017. 2

[33] U. N. News. Traffickers abusing online technology, un crime prevention agency warns. `https://news.un.org/en/story/2021/10/1104392`, 2021. 2

[34] O. O'Connell. Manhattan to stop prosecuting prostitution, dismissing cases dating back decades. `https://www.independent.co.uk/news/world/americas/manhattan-prostitution-prosecution-cyrus-vance-b1835256.html`, 2021. 9

[35] R. M. O'Donnell. Challenging racist predictive policing algorithms under the equal protection clause. *NYUL Rev.*, 94:544, 2019. 9

[36] U. N. O. on Drugs and Crime. Global report on trafficking in persons. `https://www.unodc.org/documents/data-and-analysis/tip/2021/GLOTiP_2020_15jan_web.pdf`, 2020. 8

[37] Polaris. 2020 us national human trafficking hotline statistics. `https://polarisproject.org/2020-us-national-human-trafficking-hotline-statistics/` , 2021. 2

[38] R. S. Portnoff, D. Y. Huang, P. Doerfler, S. Afroz, and D. McCoy. Backpage and bitcoin: Uncovering human traffickers. In *KDD*, pp. 1595–1604. ACM, 2017. 2

[39] R. Rabbany, D. Bayani, and A. Dubrawski. Active search of connections for case building and combating human trafficking. In *KDD*, pp. 2120–2129. ACM, 2018. 2

[40] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, Sept. 1978. 3

[41] S. Rosenthal and A. K. Dey. Towards maximizing the accuracy of human-labeled sensor data. In *IUI '10*, 2010. 3

[42] L. L. Sarah N. Lynch. `https://www.reuters.com/article/us-usa-backpage-justice/sex-ads-website-backpage-shut-down-by-u-s-authorities-idUSKCN1HD2QP`, 2018. 2

[43] D. Schuler and A. Namioka. *Participatory design: Principles and practices*. CRC Press, 1993. 3

[44] SpotLight. `https://centerforimprovinginvestigations.org/human-trafficking-investigations/`. 2

[45] Y. Sun, E. Lank, and M. A. Terry. Label-and-learn: Visualizing the likelihood of machine learning classifier's success during data labeling. In *IUI*, pp. 523–534. ACM, 2017. 3

[46] E. Tong, A. Zadeh, C. Jones, and L. Morency. Combating human trafficking with multimodal deep models. In *ACL (1)*, pp. 1547–1556. Association for Computational Linguistics, 2017. 2

[47] J. Uyheng and K. M. Carley. Bots and online hate during the COVID-19 pandemic: case studies in the United States and the Philippines. *Journal of Computational Social Science*, 3(2):445–468, November 2020. doi: 10.1007/s42001-020-00087- 9

[48] C. Vajiac, A. Olligschlaeger, Y. Li, P. Nair, M.-C. Lee, N. Park, R. Rabbany, D. H. Chau, and C. Faloutsos. Trafficvis: Fighting human trafficking through visualization. *Poster, IEEE VIS*, 2021. 2

[49] J. J. van Wijk. The value of visualization. *VIS 05. IEEE Visualization, 2005.*, pp. 79–86, 2005. 2

[50] L. Wang, E. Laber, Y. Saanchi, and S. Caltagirone. Sex trafficking detection with ordinal regression neural networks. *arXiv preprint arXiv:1908.05434*, 2019. 2, 3

[51] L. Zhang, Y. Tong, and Q. Ji. Active image labeling and its application to facial action labeling. In *ECCV*, 2008. 2