Backpage and Bitcoin: Uncovering Human Traffickers

Rebecca S. Portnoff
UC Berkeley
rsportnoff@cs.berkeley.edu

Danny Yuxing Huang
UC San Diego
dhuang@cs.ucsd.edu

Periwinkle Doerfler
NYU
pid207@nyu.edu

Sadia Afroz
ICSI
safroz@icsi.berkeley.edu

Damon McCoy
NYU
mccoy@nyu.edu

Abstract

Sites for online classified ads selling sex are widely used by human traffickers to support their pernicious business. The sheer quantity of ads makes manual exploration and analysis unscalable. In addition, discerning whether an ad is advertising a trafficked victim or an independent sex worker is a very difficult task. Very little concrete ground truth (i.e., ads definitively known to be posted by a trafficker) exists in this space. In this work, we develop tools and techniques that can be used separately and in conjunction to group sex ads by their true owner (and not the claimed author in the ad). Specifically, we develop a machine learning classifier that uses stylometry to distinguish between ads posted by the same vs. different authors with 90% TPR and 1% FPR. We also design a linking technique that takes advantage of leakages from the Bitcoin mempool, blockchain and sex ad site, to link a subset of sex ads to Bitcoin public wallets and transactions. Finally, we demonstrate via a 4-week proof of concept using Backpage as the sex ad site, how an analyst can use these automated approaches to potentially find human traffickers.

1 Introduction

Sex trafficking and slavery remain amongst the most grievous issues the world faces, supporting a multi-billion dollar industry that cuts across all nationalities and people groups [13]. With the advent of the Internet, many new avenues have opened up to support this pernicious business, including sites for online classified ads selling sex [11].

Although these ad sites provide a significant source of potentially incriminating data for law enforcement, monitoring these sites is unfortunately a labor-intensive task. The rate of new ads per day can reach into the thousands, depending on the website [14]. In addition, the nature of the advertising content can have a uniquely damaging psychological toll on its viewers. Picking out signs of trafficking requires domain expertise, creating an additional barrier for analytics. This problem space is made all the more difficult by the dearth of ground truth, e.g. ads known to be tied to trafficking activity vs. other consensual activity.

In conversation with our NGO and law enforcement collaborators, we have found that there is a real need for tools able to group ads by true owner. Such a tool would allow officers to confidently use timing and location information to distinguish between ads posted by women voluntarily in this industry vs. those by women and children forcibly trafficked. For example, groups of ads—posted by the same owner—that advertise multiple different women across multiple different states at a high ad output rate, is a strong indicator of trafficking. In this case, our goal is to distinguish which ads are owned by the same person or persons. This information can then be used to find traffickers, connections between pimps, or even trafficking networks.

All of the existing work in this problem space to date uses hard identifiers like phone numbers and email address links to define ownership. This is known to be unreliable (as criminal organizations regularly change their phone numbers/use burner phones, and the cost of creating a new email address is low) but is the best link currently available. In fact, most of the work in this domain has focused on understanding the online environment that supports this industry through surveys and manual analysis ([2, 11, 14]). Almost no work has been done in building tools that can automatically process and classify these ads [4].

The aim of this paper is to develop and demonstrate automatic techniques for clustering sex ads by owner [1]. We designed two such techniques. The first is a machine learning stylometry classifier that determines whether any two ads are written by the same or different author. The second is a technique that links specific ads to publicly available transaction information on Bitcoin. Using the cost of placing the ad and the time at which the ad was placed, we link a subset of ads to the Bitcoin transactions that paid for them. We then analyze those transactions to find the set of ads that were paid for by the same Bitcoin wallet, i.e., those ads that are owned by the same person. As far as we are aware, this is the first work to explore this connection between paid ads and the Bitcoin blockchain, and attempt to link specific purchases to specific transactions on the Bitcoin blockchain.

In addition to reporting our results using our stylometry classifier on test sets of sex ads labeled by hard identifier, we apply both our tools to 4 weeks of scraped sex ads from Backpage, a well known advertising site that has faced multiple accusations of involvement with trafficking [10]. We assess the differences and similarities between the set of owners found using just hard identifiers, our

KDD '17, August 13–17, 2017, Halifax, NS, Canada.

978-1-4503-4887-4/17/08…$15.00
DOI: 10.1145/3097983.3098082

[1] Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of DARPA.

stylometry model, the Bitcoin wallet, and finally all three combined. In summary, our contributions are as follows:

- We develop a stylometry classifier that distinguishes between sex ads posted by the same vs. different authors with 90% TPR and 1% FPR.
- We design a linking technique that takes advantage of leakages from the Bitcoin mempool, blockchain and sex ad site to link a subset of sex ads to Bitcoin public wallets and transactions.
- We propose two different methodologies that combine our classifier, our linking technique, and existing hard identifiers to group ads by owner.
- We evaluate our techniques on 4 weeks of scraped sex ads from Backpage, relying on the data automatically extracted using those two methodologies. We rebuild the price of each Backpage sex ad, and analyze the output of our two different methodologies.

We are working with two NGOs (Thorn and Global Emancipation Network) and one company (Marinus) who are all either currently using some subset of our tools and techniques, or are planning to work with us to incorporate our tools and techniques into their existing technology framework. Additionally, several law enforcement contacts have expressed a strong desire to deploy our tools in their own investigations once they become available.

The rest of this paper is organized as follows. Section 2 provides the necessary background for the rest of the paper. Section 3 outlines Backpage and Bitcoin, which we analyzed and used to evaluate our tools. Section 4 describes the methodology for building our stylometry classifier, covering ground truth labeling, the model we built, and validation results. Section 5 describes our linking technique. Section 6 describes our two proposed methodologies that combine our classifier, linking technique and existing hard identifiers to group ads by owner. Section 7 reports our findings when exploring the 4-weeks of scraped sex ads from Backpage, and Section 8 discusses limitations and future work. We conclude with reiteration of key contributions and findings.

2 Related Work

2.1 Sex Trafficking Online

Ecosystem Analysis. Much of the research in this area to date has focused on surveys, manual analysis and meta-studies to better understand the existing online environment which allows for and encourages the trafficking of humans for sexual service ([2, 3, 8, 11, 14]).

These surveys all found that the majority of US-based trafficking victims are advertised online. In Bouché's survey of 111 sex trafficking survivors ([3]), 63% of participants reported being advertised online. Of those, almost half reported that they were advertised on Backpage; Craigslist and Facebook were the other most popular websites for advertising. Latonero et al. [11] outlines several criminal cases and news stories of traffickers using online classified sites such as Backpage to sell their victims ([6, 17, 18]). In one chilling case from 2010, New York gang members reportedly advertised girls as young as 15 on Backpage, beating and starving them if they did not make at least $500 a day performing sexual services [16].

Classifying Sex Ads. Automatic analysis in the sex trafficking space is still fairly sparse, as this is an area of research just recently gaining interest in the larger computing community. What does exist has primarily focused on using machine learning to detect instances of human trafficking in escort advertisements, and using machine learning and social network analysis to detect human trafficking entities and networks in general ([4, 7]).

In [4], Dubrawski et al. present a bag-of-words machine learning model to identify escort ads that likely involve human trafficking. Using phone numbers of known traffickers as ground truth, with a false positive rate of 1% they achieve a true positive rate of 55%. They also present an entity resolution logistic regression model to group ads authored by the same person, or advertising the same person/group of people. Using personal features (age, race, physical characteristics) and operational characteristics (locations, movement patterns) with hard identifiers as ground truth, the authors conducted a small empirical evaluation with a balanced test set of 500 pairs of ads, achieving a 79% true positive rate at a false positive rate of 1%. They were also able to demonstrate some stand-alone cases where their model successfully tracked one author's ad record over the course of a year, even with phone number and a few other characteristics changing between advertisements.

Ibanez and Suthers [7] analyze Backpage sex ads using semi-automated social network analysis to detect human trafficking networks going into, and operating within, the state of Hawaii. In order to focus their attention on ads indicating trafficking, the authors first analyzed these ads for signs of trafficking derived from a list of indicators produced by the United Nations Office on Drugs and Crime and the Polaris Project. 82% of the ads contained one or more indicators and 26% contained three or more indicators. They then built a graph representing the movement of these escorts by extracting the state of origin for phone numbers listed in the ad (using the area code), and the various locations where the ad was listed. 208 total phone numbers were analyzed, and of those, 165 indicated movement. From that set, the authors discovered a potential trafficking network going from Portland, Oregon to Hawaii, as well as smaller trafficking networks within Hawaii proper.

2.2 Bitcoin

Bitcoin is a decentralized peer-to-peer pseudonymous payment system where users can transfer bitcoins among one another in the form of *transactions* that exchange this digital currency [2]. Succinctly described, a bitcoin is owned by a public encryption key, typically called a *wallet* or *address*. Transactions in practice are performed using the ECDSA signature scheme, where the owner of a bitcoin signs a statement agreeing to transfer ownership of bitcoin to another wallet (i.e. public key). Bitcoin is pseudonymous in that all transactions from a single wallet are linked to the same owner, but the same person can use many different wallets and these transactions will not be directly linkable. There have been many prior studies that point out limitations in the pseudonymous property of Bitcoin, when used in practice, and present methods for linking chains of bitcoin transactions from different wallets to the same owner [1, 12, 15]. In this study, we leverage some of these bitcoin linking methods.

[2] The standard is to use Bitcoin to refer to the system and bitcoin or BTC to refer to the digital currency.

As there is no central authority, senders first broadcast their transactions across the Bitcoin peer-to-peer network, which consists of individual volunteer nodes that each maintain the full state of the network. Upon receiving the transactions, each node stores them into a temporary storage area known as the "mempool". Transactions in the mempool *may* be selected for *mining*, a process that is meant to secure bitcoin transactions and ensure the integrity of the distributed ledger (i.e. *blockchain*). There is no guarantee whether and when a transaction will be included into the blockchain, but if this happens, the transaction will be removed from the temporary mempool and included in the permanent blockchain.

Many merchants will wait until a valid bitcoin transaction is included in the blockchain before considering it completed, which will take on average 5 minutes to occur. However, the delay time between when a client broadcasts a bitcoin transaction over the Bitcoin peer-to-peer network and when it is included in the blockchain is variable and can take hours when the network is overloaded. Due to this delay some merchants, such as Backpage, choose to not wait and accept the payment as completed once a valid bitcoin transaction appears in their mempool. By not waiting the merchant is accepting the risk of the customer performing a double spend attack [9] that causes the transaction to the merchant to be invalidated before it completes. To the best of our knowledge, there is no prior method for linking an online ad posting to the bitcoin transaction that paid for the online ad.

3 Datasets

3.1 Backpage

In this work, we focus our analysis and case study on data from Backpage, one of the most popular sites for online classified ads selling sex [11]. Backpage is widely known to be a popular domain used by traffickers to advertise their victims ([3] [11] [14]). Ernie Allen, president and CEO of the National Center for Missing and Exploited Children, makes the danger clear: "[O]nline classified ads make it possible to pimp these kids to prospective customers with little risk [for the pimp]." [16] Obviously, in order to protect the pimp/trafficker, none of these ads explicitly state any coercion; the ads are either written as if from the perspective of the victim herself, or describing the victim being sold in the third person, with no mention of a pimp or trafficker in either case.

Backpage has been running since 2004, with listings all over the world. Any person who has an email address can register for an account on Backpage and post ads. Although the site offers a wide variety of different types of classified listings (e.g., automative, rentals, furniture), in this work we focus our attention on the "adult entertainment" listings, which contain about 80% percent of the U.S. market for online sex ads in America [10]. There are several different sub-categories in this section, namely: escorts, body rubs, strippers/strip clubs, dom/fetish, trans, male escorts, phone/websites, and adult jobs. On July 1, 2015, Visa and MasterCard stopped processing transactions for adult listings on Backpage, which caused Backpage to switch to Bitcoin payments for all paid adult ads. GoCoin, a third-party Bitcoin payment processor company, currently manages all Bitcoin payments for Backpage adult ads. As of January 9, 2017, the adult listings section of the website has been blocked, in response to ongoing legal action against Backpage for their role in the marketing of minors. All of our data was collected before that point.

We have two different forms of access to this data. First, we have a scrape containing 1,164,663 unique ads from January 2008 to September 2014. We define "author" to be an entity tied to a set of hard identifiers that co-occur in any given ad. By processing all the ads and linking together phone numbers and email addresses, we discerned that we have 336,315 authors in this dataset. This data was used to build and assess our authorship classifier. Second, we conducted a scrape from December 11, 2016 to January 9, 2017, collecting all adult ads placed in the United States every hour. This data was used in our case study. Using the same definition of authorship as above, this scrape contains a total of 741,275 unique ads and 141,056 authors.

Dates	No. Unique Ads	No. Authors	Locations
1/2008-9/2014	1,164,663	336,315	Global
12/2016-1/2017	741,275	141,056	United States

Table 1: Backpage

3.2 Bitcoin

A registered user can post Backpage ads for free, but premium features, such as posting a single ad across multiple locations or bumping an ad to the top of a listings page, will require payment. For adult entertainment ads, bitcoin or a hand mailed check are the only acceptable forms of payment. GoCoin processes all bitcoin payments on Backpage.

Each purchase of premium features, however many, is represented as a single invoice. Users also have the option to deposit an arbitrary amount of bitcoins as credits; each purchase paid for via credit would withdraw funds from those pre-deposited credits. For each invoice, Backpage dynamically generates a fresh wallet address that belongs to GoCoin, along with the bitcoin amount. A user can either transfer bitcoins from his own personal wallet address into the fresh address, or he can use a third-party wallet service such as Paxful. Bitcoins received by the fresh address are subsequently aggregated into some central wallet address of GoCoin, along with bitcoins received by fresh addresses for other users.

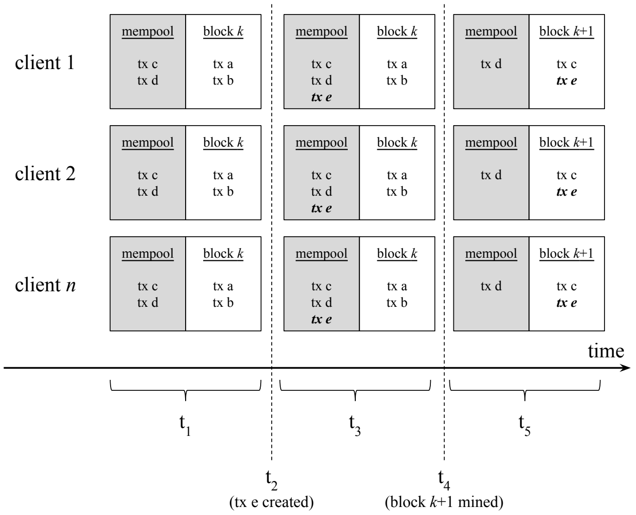
When a user transfers bitcoins into the fresh wallet address, the corresponding transaction typically appears on the Bitcoin peer-to-peer network within seconds. Once Backpage sees the transaction on their mempool, the premium features take effect and the ad appears on the listings page, without the user having to wait for the transaction to be confirmed into the blockchain. For example, if a user purchases a premium feature that posts an ad across multiple locations, the timestamp at which the ad appears across multiple locations is approximately the timestamp at which the transaction is propagated on the Bitcoin network.

This timing proximity allows us to link Bitcoin transactions, as they first appear on the peer-to-peer network, with Backpage ads. Before we can establish such links, however, we need to know exactly when a transaction first appears on Bitcoin's peer-to-peer network. To this end, we build a tool that snapshots the state of the network at a fine granularity.

In particular, we run the default Bitcoin client at our research institution. The client maintains the up-to-date blockchain, and it also allows us to query the mempool state via the `getrawmempool` API call. The mempool state is dynamic; new transactions are broadcast over the Bitcoin peer-to-peer network, while some existing transactions are removed from the mempool as they are confirmed into blocks. To this end, we set up an automated script that saves a snapshot of the mempool state every minute. Using these per-minute snapshots along with the timestamps of the snapshots, we can find the earliest timestamp at which our Bitcoin client received a transaction. These timestamps, which we call *mempool timestamps*, estimate the first time a transaction appears on the Bitcoin network. Since our Bitcoin client runs on a low-latency gigabit research network, we assume the mempool timestamps are a reasonable approximation for the true timestamps at which the transactions were sent.

To illustrate how we use the methodology above to link the timestamps of Bitcoin transactions and Backpage ads, we consider a hypothetical example as shown in Figure 1, which depicts a peer-to-peer network of n Bitcoin clients. Each of the clients maintain two pieces of state: the mempool and the blockchain. Time t_1 shows a snapshot of the n nodes. All of them currently have block k confirmed, which includes transactions ("tx") a and b. At the same time, all the n clients have both transactions c and d held in the mempool, waiting to be confirmed into the next block $k+1$.

Figure 1: Example of how a new transaction is added to Bitcoin's peer-to-peer network.



At time t_2, let us assume that someone purchased an escort ad using transaction e. This new transaction is spread across Bitcoin's peer-to-peer network. At time t_3, which we assume is a few seconds after t_2, transaction e appears in the mempools of all the clients in the network. Since we maintain a Bitcoin client ourselves and we snapshot its mempool every minute, we are likely to detect the presence of transaction e also in t_3. Here, t_3 is the mempool timestamp for transaction e. Backpage is also likely to detect the presence of transaction e, and typically will post the corresponding ad within a minute.

At this point, however, transaction e remains unconfirmed, as it is in the mempool rather than the blockchain. Let us say that at time t_4, block $k+1$ is mined, and the miner of the block decides to include transaction e. Subsequently, at time t_5, transaction e is removed from the mempool and added to the global blockchain (as is transaction c, for the same reason). Because the transaction is confirmed, we can now use Chainalysis to identify if this transaction e sent bitcoins to GoCoin (explained later in this section).

We continuously snapshot the mempool state from Oct 24, 2016 to Jan 20, 2017 and obtained 16,767,921 transactions that were later confirmed into the blockchain. Not all the transactions are relevant to our analysis. We focus on transactions that are likely to have sent bitcoins to GoCoin. We use two methods to identify transactions to GoCoin.

(1) Chainalysis Labels. Chainalysis is a private company that clusters and labels identities on the blockchain. In particular, it repeatedly deposits bitcoins into and from GoCoin, so that Chainalysis can obtain a list of fresh Bitcoin wallet addresses generated for every deposit it makes. Even though these wallet addresses are specific to one user, eventually the bitcoins from them are transferred, along with other user deposits, to GoCoin's central wallets. Since other users' deposit wallet addresses appear in the same transaction inputs as Chainalysis' wallet addresses, Chainalysis can cluster all these addresses together and label them as GoCoin. In this way, Chainalysis is able to discover wallet addresses used for making payments to GoCoin. Through its subscriber-only API, we can check if a particular transaction made payments to GoCoin.

(2) GoCoin Heuristics. While Chainalysis' technique provides us with the ground truth for transactions to GoCoin, it is unable to discover *all* GoCoin transactions. To account for false negatives, we develop heuristics to identify possible GoCoin transactions. By analyzing many GoCoin transactions that Chainalysis identified, we found that GoCoin transactions have the following unique features: (i) the fresh wallet address appeared in exactly two transactions—one for receiving bitcoins from the user, and the other for sending the bitcoins into some aggregation wallet address; (ii) the deposited bitcoin amount is always less than 1 BTC and has between 3 and 4 decimal places (e.g. "0.0075 BTC"); (iii) the bitcoins are aggregated along with other bitcoins that follow Feature (ii); and (iv) all these bitcoins are aggregated into a single multi-signature wallet address that starts with the number "3". We label any wallet address that meets all four conditions as GoCoin-heuristic. We note that this technique may introduce false positives—i.e., transactions that resemble GoCoin transactions but in reality are not GoCoin.

During the period when we snapshot the mempool state, we labeled 753,929 distinct wallet addresses as either Chainalysis-GoCoin or Heuristic-GoCoin. Of these addresses, 1.5% are Chainalysis-GoCoin only, 69.6% are Heuristic-GoCoin only, and 29.0% have both labels. For wallet addresses labelled as Chainalysis-GoCoin, 95.2% of them are also labelled as Heuristic-GoCoin.

4 Author Classifier

For any given two ads appearing on a site, we extract the authorship similarity, determining whether the ads are written by the same or different author. We take a supervised learning approach, labeling a randomly sampled subset of the data with ground truth and using those annotations to train our classifier to label the rest. We build a binary classifier that takes a pair of ads as inputs and outputs 'same' if the ads are written by the same author or 'different' otherwise.

4.1 Labeling Ground Truth

Our ground truth labeling uses hard identifiers (phone numbers and email addresses) to define ground truth authorship. Any set of co-occurring phone numbers/email addresses within the full set of ads is considered a unique author. We labeled all the data available to us in this way.

4.2 Models

We consider two models for authorship classification. For both models, we experimented with using multiple different machine learning algorithms for training. We achieved best performance using logistic regression and report only those results. We train the logistic model by coordinate descent on the primal form of the objective [5] with ℓ_2-regularization.

WritePrints Limited. This model uses a limited section of the Writeprints [19] feature set,[3] consisting mainly of counts of characters, words, punctuation, etc. This feature set has been widely used for authorship attribution. We consider the WritePrints feature set to be appropriate for our domain because the ads are similar to other short texts, such as tweets and product reviews, where this feature set has been used successfully. For each pair of ads, we extract the Writeprints limited feature set for each ad, resulting in two feature vectors. We obtain the final feature values by subtracting these two vectors and taking the absolute value of each coordinate. We use this model as a baseline for evaluating the performance of our Jaccard and Structure model.

Jaccard and Structure. This model uses a variety of text-based features: word unigrams, word bigrams, character n-grams, parts of speech, and proper names, as well as a structural feature: the location and spacing of line breaks in the post. For each pair of ads, we extract the relevant set (e.g., all adjectives) appearing in each ad, and calculate the Jaccard (on the text-based values) and cosine (on the structural values) similarity of the two sets.

4.3 Validation Results

We assessed our classifier on strictly non-overlapping sets of authors between the training and testing datasets, in order to ensure that the classifier was learning the concept of 'same' vs. 'different', and not just learning the stylometry for the particular set of authors. Ultimately, we built three separate training/testing datasets.

Building Validation Sets. In our initial pre-process, we removed all ads that were exact duplicates of each other (leaving only one copy of each duplicate in the final set) as well as all ads that had fewer than 50 words in the ad. From this set, we randomly sampled 5,000 authors with at least two ads each. From each of these authors, we randomly sampled two ads. The resulting 10,000 ads were used to create 2,500 'same' instances (where each same instance represents two ads written by the same author). We then randomly sampled 5,000 pairs of authors, where the two authors in each pair are distinct from each other, from the original sample of 5,000 authors. We then randomly sampled one ad from each of the authors in a given pair. These two ads (one for each author in a pair) were used to create 5,000 'different' instances (where each different instance represents two ads written by different authors).

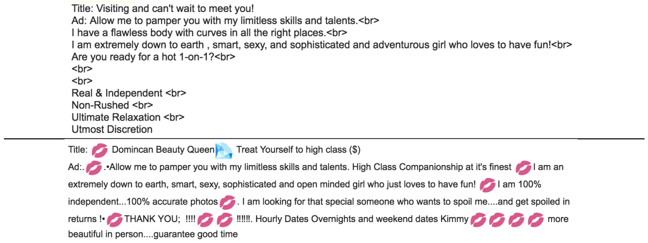
Model	TPR	FPR
Jaccard & Structure	89.54%	1.13%
Writeprints Limited	83.06%	16.93%

Table 2: Classification accuracy for same vs. different author.

We repeated this process three times, with non-overlapping sets of 5,000 authors. In this way, we created three separate training/testing datasets, with each one consisting of 7,500 instances: 5,000 different instances and 2,500 same instances. We chose this class balance in order to reflect the underlying nature of the data (i.e., there are more ad pairs with different authors than the same author). We evaluate our tool with all six different training/testing combinations, training on one of the datasets and separately testing on the other two, for all pair-wise combinations of the three datasets.

Results. In all cases, the same vs. different author classifier is effective, achieving 89.54% true positive rate and 1.13% false positive rate on average. This indicates that the classifier is not just learning to distinguish ads written by a specific set of authors, but is learning the concept of same vs. different in general. This is necessary for this domain of sex trafficking, where new victims are recruited daily, and there is no guarantee of a permanent set of traffickers persisting through time. In addition, the classifier significantly outperforms the baseline Writeprints Limited model; the accuracy for the same author class improves slightly, and the accuracy for the different author class improves dramatically in all cases.

Figure 2: Example of a false positive case with possibly flawed ground truth.



Title: Visiting and can't wait to meet you!
Ad: Allow me to pamper you with my limitless skills and talents.
I have a flawless body with curves in all the right places.
I am extremely down to earth , smart, sexy, and sophisticated and adventurous girl who loves to have fun!
Are you ready for a hot 1-on-1?

Real & Independent
Non-Rushed
Ultimate Relaxation
Utmost Discretion

Title: 💋 Domincan Beauty Queen💎 Treat Yourself to high class ($)
Ad:.💋.•Allow me to pamper you with my limitless skills and talents. High Class Companionship at it's finest 💋I am an extremely down to earth, smart, sexy, sophisticated and open minded girl who just loves to have fun! 💋I am 100% independent...100% accurate photos💋. I am looking for that special someone who wants to spoil me....and get spoiled in returns !•💋THANK YOU; !!!!💋💋!!!!!. Hourly Dates Overnights and weekend dates Kimmy💋💋💋💋 more beautiful in person....guarantee good time

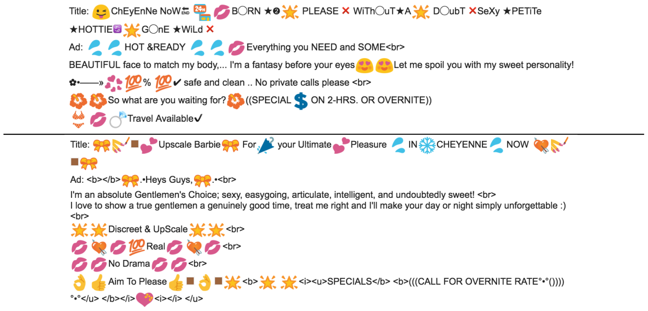
We reviewed the false positive and false negative cases from our authorship classifier. Since we use phone numbers and email addresses as ground truth, it is possible (and in some cases, appears to be the case) that the false positives are actually true positives. Figure 2 shows one such case, where the ad posters used different phone numbers and different formatting to present the exact same textual content; our authorship classifier considered them to be written by the same author. It is possible one author randomly selected, copied and pasted the text from the other, and that there is no shared owner; given the lack of definitive ground truth it is not possible to know for sure. For false negative cases, we found that the classifier misidentifies ad pairs where the writing style is completely different (Figure 3).

5 Linking Ads to Bitcoin Transactions

This section describes our method for linking a particular bitcoin transaction to its corresponding ad. Suppose A is the set of ads on the target ad site S where an individual ad $a \in A$. T is the set of bitcoin transactions whose output wallet belongs to the Bitcoin payment processor for site S. We construct an undirected bipartite

[3] We do not use the full set of Writeprints features, since it is too computationally expensive to run on larger sets of pairs.

Figure 3: Example of a false negative case where the ads appear in different sections.



Title: 😜 ChEyEnNe NoW🔚 🏪 💋 B◯RN ★❷🌟 PLEASE ❌ WiTh◯uT★A🌟 D◯ubT ❌SeXy ★PETiTe ★HOTTIE♋🌟G◯nE ★WiLd ❌

Ad: 💦 💦 HOT &READY 💦 💦 💋 Everything you NEED and SOME

BEAUTIFUL face to match my body,... I'm a fantasy before your eyes😍 😍 Let me spoil you with my sweet personality!

✿•——» 💞 💯 % 💯 ✔ safe and clean .. No private calls please

🌺 🌺 So what are you waiting for? 🌺 ((SPECIAL 💲 ON 2-HRS. OR OVERNITE))

👙 💋 💍 Travel Available✓

Title: 🎀 💅 ■ 💕 Upscale Barbie 🎀 For ☔ your Ultimate 💕 Pleasure 💦 IN ❄ CHEYENNE 💦 NOW 💘 💅 ■ 🎀

Ad: <b></b> 🎀 .•Heys Guys, 🎀 .•

I'm an absolute Gentlemen's Choice; sexy, easygoing, articulate, intelligent, and undoubtedly sweet!
I love to show a true gentlemen a genuinely good time, treat me right and I'll make your day or night simply unforgettable :)

🌟 🌟 Discreet & UpScale 🌟 🌟

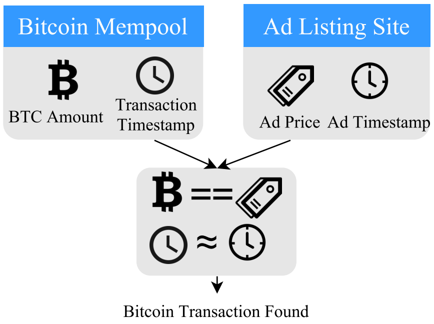
💋 💘 💋 💯 Real 💋 💘 💋

💋 💋 No Drama 💋 💋

👌 👍 Aim To Please 👍 ■ 👌 ■ 🌟 <b> 🌟 🌟 <i><u>SPECIALS</b> <b>(((CALL FOR OVERNITE RATE°•°()))) °•°</u> </b></i> 💝 <i></i> </u>

graph, $G = (V, E)$, where the set of vertices is $V = A \cup T$, and the set of edges E contains an edge between two nodes, $a \in A$ and $t \in T$, if t is a possible transaction for a. We consider t to be a possible transaction for a if the cost of posting a equals the value of t, and the difference between the timestamp of a's appearance on the listings page and the timestamp when t is first observed on either the mempool or the blockchain is less than a threshold. The threshold depends on the particular ad site. For example, the threshold for Backpage is one minute. Backpage accepts the payment for an ad as completed, and posts said ad on S, as soon as t appears in their mempool. We then observe that transaction in our mempool within one minute. In this case, the value of the threshold is simply the amount of time it takes for the transaction t to appear in our mempool: one minute.

Figure 4: Linking Ads to Bitcoin Transactions



The edge between transaction t and ad a could be a false positive if the transaction, in reality, is not linked to the ad. For example, if a was not paid for in bitcoin, and there happened to be a transaction t around the same time with the same cost, that would result in a false positive. False negatives, where there is a missing edge between t and a, are also possible. False negatives can occur if we wrongly reconstructed the price of a and thus failed to link a to the corresponding t.

6 Grouping Ads by Owner

We propose two methods for grouping ads by owner: grouping by shared hard identifiers and grouping by persistent Bitcoin identities. Our methods assume two sources of data: the cost to post each ad a in A, and the timestamp of a's appearance on the target site. The mechanism for collecting this timestamp data and rebuilding the cost varies from site to site. In section 7, we demonstrate how we did both for Backpage.

Both methods have T to be a set of bitcoin transactions to GoCoin, such that each transaction $t \in T$ has exactly one input wallet address w that is not multi-signature, and at least one of the output wallet addresses in t is either labeled as Chainalysis-GoCoin or Heuristic-GoCoin. The ultimate goal is to map an ad a to its true owner wallet address w, from among all W (the set of wallet addresses on the blockchain). Our first method tackles this problem by finding a wallet address w that links together multiple existing hard identifiers. Our second method focuses on persistent Bitcoin identities (defined below).

Figure 5: Shared Hard Identifiers

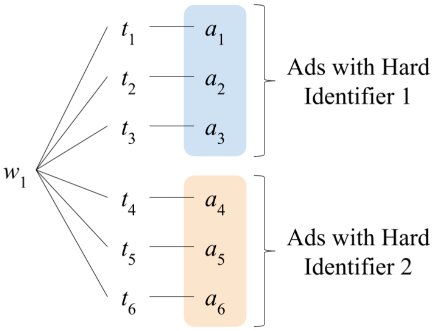
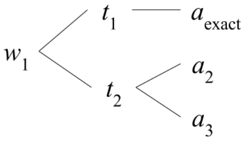


Figure 6: Persistent Bitcoin Identities



6.1 Grouping by Shared Hard Identifiers

In practice, the mapping between a transaction t and an ad a is many-to-many. This many-to-many mapping between t and a makes it difficult to map a wallet address w to a. To this end, we construct a subgraph $G' \subset G$, where G is an undirected graph with A, T, and W as the vertices. In G, an edge between a w node and a t node exists if wallet w is the sole input wallet address of transaction t. We require subgraph G' to satisfy all of the following criteria.

1. Each t node should be adjacent to exactly one w node, because we already require every transaction in G to have a single input wallet address. However, we allow each w to be adjacent to one or more t, as a wallet address may be used across multiple transactions.
2. With exactly two hops (from w to t to a), each w node should be able to reach at least three a nodes with the same hard identifier. This reachability suggests that w is likely to be the true owner for at least three of the a nodes; the presence of the shared hard identifier reduces the probability of having incorrect edges between t and a.
3. Each t should be adjacent to exactly one a. By transitivity, each a can reach exactly one w with two hops. In other words, one cannot find another wallet address, other than w, that can be mapped to ad a. This criteria attempts to further reduce the probability of incorrect edges between t and a.

In an effort to link together multiple hard identifiers, we add one more criteria:

(1) With exactly two hops, each w should be able to reach at least two sets of a nodes with different hard identifiers. This suggests that these hard identifiers are likely to be related, in that they might have all used w to pay for the ads.

We define the resulting w in G' to be a `shared hard identifier` wallet (SHI wallet). The last criteria allows us to find at least two sets of ads of different hard identifiers that are mapped to the same w. Even so, because of the false positive/false negative problem we are unable to definitively conclude that w was used to pay for the ads. Further investigation and some manual analysis is needed to establish that the ads are indeed related, despite having different hard identifiers.

Figure 5 shows a hypothetical example of a subgraph G' that satisfies all our criteria. In particular, the wallet address w_1 is associated with two groups of transactions and ads that are linked to two distinct hard identifiers. Within each group, there is a one-to-one mapping between each transaction and ad; for example, transaction t_i is linked to ad a_i for $i = 1, 2, \cdot, 6$. In this way, w_1 is the shared hard identifier for these six ads.

6.2 Grouping by Persistent Bitcoin Identities

Within the set of input wallet addresses for transactions $t \in T$, any w that sends the change from each of its transactions t back into w is a `persistent bitcoin identity` (PBI). We further reduce this set of PBI's by only keeping those where at least one t is adjacent to exactly one a, and that ad a is also adjacent to only that t – i.e., they are an 'exact match.' When an exact match is found, we consider the PBI to be the owner of the matching ad. Figure 6 shows a hypothetical example of wallet address w_1 as a PBI for ad a_{exact}, as there is a one-to-one mapping between t_1 and a_{exact}.

For any remaining transactions that are not exact matches, e.g. multiple a's are linked to the same t, we use the author classifier (Section 4) to find the most likely true link. In our hypothetical example in Figure 6, t_2 fits that category, linked to both a_2 and a_3. For each such a_i that is linked to the transaction t, we run our binary author model on each pair a_{exact}, a_i. The ad a_i that belongs to the a_{exact}, a_i pairing with the highest probability of being written by the same author is selected as the matching ad.

7 Case Study

To validate our methods, we placed 33 Backpage ads – 32 paid, and one free – and used these as ground truth. Eleven of these were paid for using our personal `1Ejb3` persistent wallet address. We focus our validation on these eleven, as both our methods of grouping ads by owner focus on wallets that get reused for multiple transactions. The ads were placed from Dec 12, 2016 until Dec 24, 2016. The price of the ads ranged from $2-$20. For all eleven `1Ejb3` ads the main contributor of the cost was bumps (see Section 7.1 for more details). Over 87% (29) of the ads were in Escorts category. The ads were posted in 27 distinct US states and regions.

7.1 Price Reconstruction

To reconstruct the price of an ad we reverse engineer the exact algorithm used by Backpage. To that end, we placed several ads ourselves, studied the html of the price payment page, and reproduced the algorithm (see Algorithm 1). To calculate the price of an ad, we need to know how often an ad was 'bumped' (e.g., appeared on the main listing page exactly an hour later), how often an ad was 'reposted' (e.g., appeared on the main listings page in the same hour every day for X number of days, where the number of days must be in 4-day increments), how many weeks an ad was 'sponsored' (e.g., appeared in a thumbnail highlighted on the side of the main listings page as well as in its normal position on the body of the listings page) and how many locations to which the ad was posted.

In order to collect this information, we wrote a scraper that scraped, every hour, all of the listings pages for every region (totaling 67) in the United States, for every adult entertainment subcategory. Backpage includes a minute-granularity timestamp indicating when an ad was posted; we extracted this timestamp during the scrape to determine the first appearance of an ad, and subsequent bump, repost and sponsor appearances. We ran our scraper for 4 weeks, shutting it down on January 9th when Backpage took down its adult entertainment section. In parallel, we also ran a separate scraper that collected the pricing information for each of the 67 regions, as the cost of a bump, repost, sponsor or location varied depending on the region to which it was posted.

Based on prior scrapes, we observed that for all of the adult entertainment sub-categories except for Escorts, the price stayed constant from week to week; so we ran our pricing scraper on the non-Escorts sub-categories across all 67 regions four times, once for each week of the scrape. We observed more variability in the Escort pricing, and therefore scraped that sub-category's pricing once a day. We also noted that the sponsor pricing was significantly more variable than either bump, repost or location; the pricing followed what seemed to be a surge pattern where the prices varied every 15 minutes. We did not have the computing capacity to run our pricing scraper on all 67 regions every 15 minutes for four weeks, so instead we ran the pricing scraper at this rate for just one region, Los Angeles. Even scraping every 15 minutes did not allow us to reconstruct the price of our sponsored ads in Los Angeles with complete accuracy. In a previous experiment, we had scraped the pricing for Los Angeles as often as possible for one day (about every 8 minutes) and noted one instance where the sponsor price changed in as little as 10 minutes. Because this was rare, we elected during the 4-week scrape to reserve computing resources for other tasks.

Finally, we ran one last scraper that collected the first page of the main listings page for each region, for each adult entertainment sub-category, once a day, in order to collect the set of sponsored ads.

We found that for all non-sponsor ground truth ads, we correctly calculated the exact price for 24 of 25 total. For the wrong ad, the hourly scraper missed one bump that happened to occur right on the hour, so the price was calculated incorrectly. For the sponsor ground truth ads, the posting pattern was correctly extracted, but the price was incorrect due to the high variability of sponsor pricing, with predicted prices varying within +/-5% of the true price. As a result, we decided to not include any sponsor ads in the rest of the study, leaving 95% (143,908 of 151,482) of the paid ads available for our analysis. None of the 11 ads placed using our personal `1Ejb3` wallet address were sponsor ads; eight of the 21 ads placed using Paxful were sponsor ads.

Ad Type	Adult Jobs	Body Rubs	Datelines	Escorts	Fetish	Male Escorts	Strippers	Transsexual Escorts
Unique Ads	14,143	83,158	5,443	555,394	14,227	27,638	6,245	35,027
Unique Postings	46,160	382,843	27,495	1,805,174	43,166	55,351	16,881	159,778
Avg Num Locations	1.10	1.04	1.74	1.02	1.10	1.04	1.35	1.05

Ad Price	Adult Jobs	Body Rubs	Datelines	Escorts	Fetish	Male Escorts	Strippers	Transsexual Escorts	Total
Free	12,553	60,704	3,732	447,319	11,729	24,583	4,095	24,976	589,961
$1-5	1,029	3,991	813	13,703	1,660	2,196	1,392	5,967	30,751
$5-20	393	7,825	437	71,622	549	557	492	2,682	84,557
$20-100	157	7,331	161	16,987	269	194	261	1,250	26,610
>$100	11	3,307	298	5,763	20	8	5	152	9,564

Table 3: Distribution of Ads by Category

Algorithm 1 Price Recreation

```
procedure DETERMINE SERVICES
    ads ← sort(occurrences of this ad)
    bumps ← 0
    reposts ← 0
    sponsorWeeks ← number of weeks ad was sponsored
    lastAd ← ads.pop
    for ad in ads do:
        if ad.hour = lastAd.hour + 1 then
            bumps ← bumps + 1
        else if ad.day = lastAd.day + 1 and ad.hour = lastAd.hour then
            reposts ← reposts + 1
        else
            call Reconstruct Price
        lastAd ← ad
procedure RECONSTRUCT PRICE
    priceInfo ← price info for each location at time of ad
    totalPrice ← 0
    for price in priceInfo do:
        if priceInfo.size > 1 then
            $totalPrice \leftarrow totalPrice + price.base$
        $totalPrice \leftarrow totalPrice + (price.getBump * bumps)$
        $totalPrice \leftarrow totalPrice + (price.getRepost * reposts)$
        $totalPrice \leftarrow totalPrice + (price.getSponsor * sponsorWeeks)$
    return totalPrice
    if not ads.empty then
        goto Determine Services
```

7.2 Linking Backpage Ads to Bitcoin Transactions

Before attempting to group the ads by true owner using our two methods, we first had to link the scraped, paid ads to the set of transactions T (as defined in section 6). To that end, we:

(1) Constructed the w and t vertices and edges using the blockchain dataset
(2) Construct the a vertices using the Backpage scrape dataset
(3) Constructed an edge between t and a if their timestamps were within one minute of each other (using the mempool timestamps dataset), and if the ad a’s cost was within 2% of one of t’s GoCoin output values in US Dollars

Of the 11 GoCoin ground truth transactions processed with our personal 1Ejb3 wallet address, eight were an exact match for the correct ground truth ad. All three of the remaining transactions matched to two ads, one of which was the correct ground truth ad. Overall, of the 54,799 transactions in the set T found during the course of the 4-week study, 5,310 were exact matches to one ad, 47,785 matched multiple ads, and 1,704 were a match to one ad, where that ad matched to multiple transactions.

7.3 Results

7.3.1 Using Shared Hard Identifiers Using the methodology in Section 6.1, we constructed the graph G and subsequently the subgraph G'. By applying the four criteria in Section 6.1, we found 30 SHI wallets.

Each of the SHI wallets mapped to multiple ads with different hard identifiers. In order to verify whether any of those hard identifiers we had found were actually linked, we manually analyzed the title of every ad in our scrape associated with each hard identifier. If any two ads from two different hard identifiers shared the exact same title, we classified those two hard identifiers as having a true link.

On the upper end, the SHI wallet 1LYE mapped to 64 ads across 17 sets, where within each set all of the ads share the same hard identifier. After the manual analysis we found that only two of the 17 sets of ads are related. The ads from both of these sets are advertising Asian and Latina women from Los Angeles in the Escorts section, while the remaining sets range from advertising massages in Los Angeles to ‘two-girl specials’ in Massachusetts, New Jersey and New York. These results suggest that the remaining 15 sets of ads are false positives for the SHI wallet 1LYE.

On the other hand, there are SHI wallets like 1Abg and 1yVF that appear to have zero false positives. Each SHI wallet is mapped to two sets of ads. For 1Abg, both sets advertise young Asian women and specify a policy of ‘outcall only’ (e.g., the escort is the one who travels to the client). One set advertises in the SF Bay area, the other in Illinois. For wallet 1yVF, both sets again advertise young asian women: one in the SF Bay and the other in Colorado.

For most of the SHI wallets, there is at least one false positive. In the SHI wallet 1BT6, for instance, ads are mapped to 8 sets of same-author ads; 5 of those sets all advertise college-aged women from Hong Kong, Malaysia and Taiwan. This is a case where false positives are a minority. An additional example is 1Mhe, which is mapped to 6 sets of same-author ads; 4 of those sets all advertised women who were new to the United States. In contrast, there are cases where false positives are a majority, such as 1N7V, which is mapped to 8 sets of same-author ads, but we find only 3 sets to be related. In total, out of the 30 SHI wallets, false positives are a majority in 20 wallets. From across all of these SHI wallets, we are able to extract 15 new 'owner' identities, each of which are made up of two or more hard identifiers.

7.3.2 Using Persistent Bitcoin Identities We found 249 PBIs that meet the conditions set in Section 6.2. Of those, there were 90 wallets with at least one exact match transaction. Of those, 47 had at least two exact match transactions. Looking at that set of 47 persistent identities, we were surprised to observe that only two wallets (one of which was the wallet we used to make ground truth purchases) had the same hard identifier label across the set of exact match transactions. In fact, upon further investigation of the other 45 wallets and their exact match transactions, we could not find any consistent connection between the exact match transactions within a particular wallet: not hard identifier label, not text similarity, not location, and not adult service.

The single wallet `1MDJ` we observed (besides our own) that did have a shared hard identifier across multiple exact match transactions worked perfectly under our exact match method. Out of `1MDJ`'s seven total transactions in the timespan of our scrape, four were exact matches. For two of the remaining three transactions, the transaction matches to multiple ads. However, for those two transactions, exactly one of the matching ads shares the same hard identifier as the four exact match transactions. When tested against the pairs of ads, our stylometry model returns the correct match in both cases. For the final transaction, the transaction matches to only one ad, which shares the same hard identifier as the exact match transactions. However, that ad also matches to other transactions not made by the same wallet `1MDJ`. Our stylometry model returns that this ad matches to the transaction made by wallet `1MDJ`, and not to the other transactions. Every single ad placed by the hard identifier is accounted for by transactions made by the wallet `1MDJ`. Similarly, our own `1Ejb3` wallet also worked perfectly under our exact match method (with eight exact match transactions, and three multiple-match transactions where our stylometry classifier returned the correct transaction match in all three cases).

Given the seeming contradiction in exact match transactions for the other wallets, we loosened our requirements to look for wallets that were not exactly PBIs (because they did not send the change back to their own wallet with each GoCoin transaction), but did have repeat transactions to GoCoin from their particular wallet, at least two of which were exact match transactions. We found four wallets that satisfied this criteria: `16qB`, `1L8r`, `1N7A` and `1H5t`. All four of them worked perfectly under our exact match method, with every single ad placed by the relevant hard identifier accounted for by the transactions made by the corresponding wallet.

7.3.3 Bitcoin-based Owners vs Hard Identifiers Given the set of 15 new Bitcoin-based owner identities we extracted using the shared hard identifiers method, we assess what new information we learn from the grouping of hard identifiers into sets, that we would not have been able to know otherwise.

On a whole, the new Bitcoin-based owner identities consistently expanded the set of locations; what previously was a single hard identifier in a single location becomes a network of hard identifiers across multiple locations. In addition, hard identifiers that previously looked fairly small and financially limited suddenly boom out when linked to a different hard identifier that has much more capital (as judged by the average price of an ad they purchase).

One apt example is with the Bitcoin-based owner extracted from SHI wallet `1BT6`. Taken separately, each of the five hard identifiers that make up the owner is quite active. Four are based in SF Bay while the last is in Northeastern Texas. All five posted at least twice as many paid ads as free ads; each ad cost on average between $80 and $130, depending on the hard identifier. What previously looked like five distinct businesses suddenly becomes more suspicious, with the added location and combined financial resources.

8 Discussion

Both grouping by shared hard identifier and grouping by persistent bitcoin identity have an obvious limitation: false positives and negatives on the link between transactions and ads. An exact match transaction is not necessarily correct simply by virtue of being an exact match, and just because multiple ads that share the same hard identifier all map to the same wallet does not mean those mappings are correct.

In particular, there are numerous reasons why a transaction might be an exact match with an ad, but that pairing still be incorrect. The post listings scraper may have scraped an ad right on an hour boundary; the ad might have been paid for using credit; the transaction might not be a payment for an ad posting but for purchase of credit; the transaction might not be a true GoCoin transaction; the transaction might be GoCoin but not a payment for Backpage.

It is also entirely possible that the exact match transactions we found are in fact correct, and represent the presence of some third party middleman who is making Bitcoin payments on behalf of a variety of Backpage users. This seems unlikely given the organization that would be required, but is possible. This is another place where the lack of ground truth becomes quite punitive; if the results do not look correct, it is difficult to tell whether that is because it is showing us something new, or because something is wrong.

Given the correctness of all of the exact match transactions we made using our personal `1Ejb3` wallet address, and the other handful of similarly correct appearing persistent bitcoin identities we found (both with an exact match transaction and otherwise), we plan to continue parsing out of this problem in future work. There are several avenues of approach we can take in parallel. We can work to disambiguate Backpage credit payments on the Bitcoin blockchain from Backpage ad payments by analyzing ads and credit payments we make ourselves. We can target a narrow scope of ads where we are absolutely certain that every single ad's posting pattern is correct, perhaps by manually verifying by physically watching the listings page. We can show our data to law enforcement officers and work together to build a ground truth set that we can then use to validate or reject the correctness of our exact match transactions.

We are also interested in expanding the shared hard identifier method, given the promising results of finding multiple linked hard identifiers. It is important to remember the immense size of this data - the fact that we can narrow down from hundreds of thousands of ads to find these connections is remarkable, and potentially enormously helpful to those law enforcement officers who have to read through so many ads during an investigation. This both saves these officers a substantial amount of time and also protects them from some of the psychological repercussions of analyzing this data.

We also plan to continue this work past persistent Bitcoin identities. It is promising to see that all of our Paxful transactions also matched with the correct Backpage ad. There is existing research in the space of clustering Bitcoin wallets together to find groups of wallets that, while seemingly disparate, actually belong to the same owner. The ideal scenario would be one in which we used Bitcoin clustering techniques to link our Paxful transactions to each other, and then our stylometry model to tie those ads that match the Paxful transactions to the ads that match the transactions made using our persistent Bitcoin wallet. We are especially motivated to find more of these ad owners, because the results of our case study indicate that even with a small increase in linkage across hard identifiers using Bitcoin and stylometry, we can potentially find critical information (e.g., connections between previously unconnected ads that indicate movement across multiple states/geographic locations, with multiple parties involved, both of which are strong indicators of trafficking) that could help our NGO and law enforcement partners in their mission to find and rescue trafficked humans.

There is also value in simply just linking ads to transactions, even in the case where a trafficker is not reusing the same wallet. Law enforcement could potentially subpoena information from Paxful or some other wallet service; going from a set of ads of interest to the set of matching Bitcoin transactions would make it possible to then get explicit personally identifying information from a wallet service. Some of our law enforcement collaborators have also stressed the value of having the Bitcoin transaction matched to a target ad when it comes to building a case for the court, after the alleged trafficker or pimp has been arrested. Our success in matching transactions to the correct ad for some of the PBIs, SHIs, and our ground truth, is encouraging to this end.

It is worth noting that none of this work is stymied by the fact that Backpage has shut down their adult entertainment section. The vast majority of those ads shifted over to the dating section of their website, where ads are also paid for using Bitcoin. It is also worth noting that perpetrators have no choice but to continue to use Bitcoin even after our published work: the original move to Bitcoin was because of Backpage's response to Visa and MasterCard's decision to stop processing transactions for adult listings on Backpage. Perpetrators have no choice but to use the payment platform provided by the advertising company. Even if Backpage changes the virtual currency it accepts as payment, as long as that virtual currency is implemented with publicly accessible ledgers, our techniques will continue to work.

9 Conclusion

In this paper, we proposed an automated and scalable approach for identifying sex trafficking using multiple data sources. We developed a stylometry classifier and a Bitcoin transaction linking technique to group sex ads by owner. To the best of our knowledge, this is the first such work to attempt to link specific purchases to specific transactions on the Bitcoin blockchain. We evaluated our approach using real world ads scraped from Backpage, and demonstrated that our approach can group multiple ads by their real owners. We are currently collaborating with multiple NGOs and law enforcement officers to deploy our tools to help fight human trafficking. In addition to sharing our tools and data with our collaborators, we also intend to make them publicly available.

Acknowledgments

This work was supported in part by the National Science Foundation under grant CNS-1619620, by the Amazon "AWS Cloud Credits for Research", by the U.S. Department of Education, by Giant Oak, and by gifts from Google and Thorn. We thank all the people that provided us with assistance in analysis; in particular we thank Chainalysis for the use of their tools and their contribution to this work. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] Elli Androulaki, Ghassan O. Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. 2013. *Evaluating User Privacy in Bitcoin.* Springer Berlin Heidelberg, 34–51.

[2] Kristie R Blevins and Thomas J Holt. 2009. Examining the virtual subculture of johns. *Journal of Contemporary Ethnography* 38, 5 (2009), 619–648.

[3] V Bouche and others. 2015. A report on the use of technology to recruit, groom and sell domestic minor sex trafficking victims. (2015).

[4] Artur Dubrawski, Kyle Miller, Matthew Barnes, Benedikt Boecking, and Emily Kennedy. 2015. Leveraging publicly available data to discern patterns of human-trafficking activity. *Journal of Human Trafficking* 1, 1 (2015), 65–85.

[5] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A Library for Large Linear Classification. (2008), 1871–874.

[6] Federal Bureau of Investigation, Jacksonville Division. 2010. Jury Finds New York Man Guilty of Sex Trafficking Women by Force, Threats of Force, and Fraud. (2010). http://www.fbi.gov/jacksonville/press-releases/2011/ja021711.htm.

[7] Michelle Ibanez and Daniel D Suthers. 2014. Detection of domestic human trafficking indicators and movement trends using content available on open internet sources. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on.* IEEE, 1556–1565.

[8] Lara Janson, R Durchslag, and H Mann. 2013. "Our great hobby": An analysis of online networks for buyers of sex in Illinois. *Chicago Alliance Against Sexual Exploitation (CAASE)* (2013).

[9] Ghassan O. Karame, Elli Androulaki, Marc Roeschlin, Arthur Gervais, and Srdjan Čapkun. 2015. Misbehavior in Bitcoin: A Study of Double-Spending and Accountability. *ACM Trans. Inf. Syst. Secur.* 18, 1, Article 2 (May 2015), 2:1–2:32 pages.

[10] Nicholas Kristof. 2016. Every Parent's Nightmare. The New York Times. (2016). http://www.nytimes.com/2016/03/10/opinion/every-parents-nightmare.html.

[11] Mark Latonero. 2011. Human trafficking online: The role of social networking sites and online classifieds. (2011).

[12] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. 2013. A Fistful of Bitcoins: Characterizing Payments Among Men with No Names. In *Proceedings of the 2013 Conference on Internet Measurement Conference (IMC '13).* ACM, 127–140.

[13] Polaris. 2017. Human trafficking. Polaris Project. (2017). http://www.polarisproject.org/human-trafficking/.

[14] D Roe-Sepowitz, J Gallagher, L Martin, G Snyder, K Hickle, and J Smith. 2012. One-day sex trafficking snapshot of an internet service provider: Research Brief. (2012).

[15] Dorit Ron and Adi Shamir. 2013. *Quantitative Analysis of the Full Bitcoin Transaction Graph.* Springer Berlin Heidelberg, Berlin, Heidelberg, 6–24.

[16] Malika Saada Saar. 2010. Girl Slavery in America. (2010). http://www.huffingtonpost.com/malika-saada-saar/girl-slavery-in-america_b_544978.html.

[17] U.S. Department of Justice. 2011. Dallas Felon Admits to Sex Trafficking a Minor and Possessing an Assault Rifle. (2011). http://www.justice.gov/usao/txn/PressRel11/wilson_clint_ple_pr.html.

[18] U.S. Immigration and Customs Enforcement. 2009. Maryland man pleads guilty in sex trafficking conspiracy involving 3 minor girls. (2009). http://www.ice.gov/news/releases/0907/090716baltimore.htm.

[19] Rong Zheng, Jiexun Li, Hsinchun Chen, and Zan Huang. 2006. A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the American society for information science and technology* 57, 3 (2006), 378–393.