2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Detecting Sex Trafficking Circuits in the U.S. Through Analysis of Online Escort Advertisements

Michelle Ibanez
Center on the Family
University of Hawai'i
Honolulu, HI USA
ibanezm@hawaii.edu

Rich Gazan
Dept. of Computer & Information Sciences
University of Hawai'i
Honolulu, HI USA
gazan@hawaii.edu

Abstract— **The use of social network analysis is used to identify possible victims of sex trafficking and the observation of domestic trafficking flows across the U.S. A novel approach to the analysis is presented with the understanding further development is needed. Online escort advertisements were analyzed for virtual indicators of sex trafficking and to identify patterns of activity that may be associated with criminal networks. The results suggest that 67% of the sample contained indicators of movement. Social network analysis methods were applied to identify movement trends across the US. The use of SNA methods allowed prominent hubs and circuits of this activity to be observed, by providing a tool to uncover covert network structures and activity, yielding a method to capture movement trends of potential trafficked persons. The integration of geospatial data allowed maps to be created in order to visualize movement patterns.**

Keywords—Technology Facilitated Sex Trafficking, Covert Networks, Domestic Trafficking Circuits

I. Introduction

Human Trafficking is a global issue that violates human rights and is a form of modern day slavery. Current estimates indicate that as many as 27 million people are victims of various forms of human trafficking [1]. Boyd [2] developed a framework identifying 15 aspects of trafficking being reshaped by technology. This study will focus on two of those aspects: the Advertising and Selling of Victims, and Searching for and Purchasing of Victims by "Johns". Much of the advertising and searching activity occurs on online classified sites, yet there is little research examining the use of online classifieds in sex trafficking [3]. As part of a larger study [4], a sample of Backpage online escort advertisements from three U.S. cities was analyzed for indicators of sex trafficking, with a particular focus on 'Movement' indicators to identify sex trafficking circuits within the U.S. Social network analysis was the primary method used to identify and construct networks of movement trends, and geographic data was incorporated in order to visualize movement patterns.

II. Background

Technology has been cited as a significant facilitator of human trafficking [2,3,5,6,7]. Two key technologies being used to advance this market are the Internet and mobile phones, which make detection difficult and reduces risks to both buyers and sellers [6]. On the supply side, traffickers use the Internet to more efficiently reach a larger customer base and advertise the sale of their product. It also provides a media-rich environment ideal for advertising at a lower cost than other forms of advertisement [5]. On the demand side, the Internet provides clients with an anonymous environment to connect with like-minded individuals, and access advertisements and profiles of women and/or children, reducing the risk of exposure. The primary research question driving this study is: How can content available in online escort advertisements be used to observe movement trends of potential traffickers or victims?

IEEE/ACM ASONAM 2016, August 18-21 2016, San Francisco, CA, USA
978-1-5090-2846-7/16/$31.00 ©2016 IEEE

A. Use of Mobile Phones

Criminal networks rely heavily on mobile connectivity for operational and coordination purposes, typically using prepaid mobile phones, so they cannot be linked to a specific individual through a service contract [6]. Critically, these phone numbers can be used as persistent, unique identifiers in online advertisements, allowing Johns to review and share information about their experiences as a reputation system, and for sex workers, traffickers or pimps to schedule and coordinate services [7]. Many escort review and search sites use phone numbers as search terms to access provider profiles.

Phone numbers can also provide movement data. The area code provides information on the origin of the phone, which may indicate where the provider or trafficker is from [7]. Also, if the advertised location is different from the location of the area code, it can indicate movement occurred between those two locations. Phone numbers are embedded in the online classifieds as providers are advertised in different cities at different times. These ads can be accessed using a phone number search, which will access advertisement history. Documenting the string of cities a person has been advertised in is one way to track the individual's circuit.

B. Movement

Movement is a critical element of sex trafficking. While it does not define trafficking, it is a key indicator of potential trafficking activity, as movement can indicate a more coordinated effort of activity. Providers are moved from city to city on a circuit in order to maximize profits and reduce the potential of detection by law enforcement. This paper examined one of the primary indicators of sex trafficking in online advertisements suggested by Ibanez [4]: Movement (circuits). A circuit is the series of cities providers are moved to and sold in, which could include movement along entire regions [13]. This systematic movement of providers to various cities across the nation for the purposes of commercial sex activity indicates a more sophisticated, organized criminal activity taking place.

Frequent movement is also a control mechanism used by traffickers to keep victims isolated. The continuous movement and short durations of stay prevent victims from establishing social support systems and limits familiarity with a location [7]. Movement trends are profit driven, with providers often being moved to areas where maximum profits are anticipated. Also, rotating providers in and out of different cities keeps the supply stream "fresh." Providers are advertised as being 'new' or 'available for a limited time only,' enticing the demand side of the market.

C. Covert Networks

A distinct feature of covert networks is the consistent need to balance secrecy with efficiency [14,15,16]. These networks constantly have to balance the need to communicate externally with members on the other side of the market while maintaining a level of discretion about their illicit nature in order to avoid detection by law enforcement. Sparrow [17] argued for the use of social network analysis (SNA) as an effective method to exploit criminal data, citing its ability to highlight network vulnerabilities and transform raw data into intelligence. SNA techniques can be used to process large volumes of data to detect hidden structures and patterns of criminal networks.

III. METHOD

A mixed methods approach was used to capture and analyze data. Backpage was selected as the data source for this study because it is the leading site hosting online classifieds for adult services. Backpage dominated the market in terms of highest volume of posts and web traffic, with an estimated 3.3 million unique visitors generating approximately $2.6 million in March 2012. The sample was harvested using web crawling software. Ads were sampled from three hub cities based on statistics reported by the Department of Justice and the National Human Trafficking Resource Center. McClain and Garrity [22] define destination or hub cities as areas with the greatest demand, typically locations near military bases, truck stops, conventions, and tourist areas. The locations selected for this study were Hawai'i; Portland, Oregon; and Miami, Florida.

In addition to online classifieds, searching activity occurs on escort review sites. Escort review sites allow Johns to validate provider authenticity by reviewing advertisement history. These sites include information on the various cities providers have been advertised in, different aliases and ages used, photos, and reviews. These sites serve the community's needs by providing a single source of consolidated information on a provider. The various sites label the data field for other advertised locations as either 'cities worked' or 'cities traveled.' The assumption is that the women have traveled to these locations based on the presence of an advertisement for service in this location. The three escort review sites used for this study were: escortphonesearch.com, escort-phone.com, and escortsincollege.com.

After collecting Backpage escort ads from three locations across the U.S., data fields were extracted from the ads and analyzed for indicators of sex trafficking. Next, provider movement data was collected using the phone numbers in the ads, and analyzed through the escort review sites. The advertised location data obtained for each unique phone number across the four sites was integrated to maximize the available data on a provider's circuit, in an effort to circumvent the deceptive nature of this covert network.

SNA methods were used to construct provider-by-location networks, which allowed prominent movement trends to be observed. Centrality measures were used to identify the most prominent locations providers were advertised in. The node list was composed of the unique phone numbers identified in the sample, which served as a proxy for the provider or trafficker. Location data was entered using other advertised locations based on the data collected in the phone number search. Because nodes represented a geographical location, geospatial data was incorporated to visualize the data, by adding latitude and longitude coordinates to the node attributes allowing for the network to be overlaid on a map of the U.S. This analysis provided information on sex trafficking 'hotspots' across the country. Hotspots are areas with high crime density [20,21]. Crime is not evenly distributed across a map; it is concentrated in some areas and absent in others.

Circuits were observed based on aggregate data using a monopartite projection of the network. Specifically, the network was folded to create a location-by-location network. Network edges represent locations sharing telephone numbers, with the edges weighted by the number of phone numbers shared between locations. Then, thresholds were established to filter data by edge weight in order to identify high volume travel routes. The resulting visualization of high volume associations between locations was used to identify prominent hubs and circuits. This level of data provides insight on critical locations in the network allowing resources to be targeted to areas with high volumes of potential sex trafficking activity.

A. Sample

Online advertisements for escort services from Backpage Hawai'i; Portland, Oregon; and Miami, Florida were collected on May 10, 2014. A total of 200 advertisements from each of the three target locations were collected (n=600). There was a total of 603 unique providers identified in the sample (Female=602, Male=1). The advertised age ranged from 18-

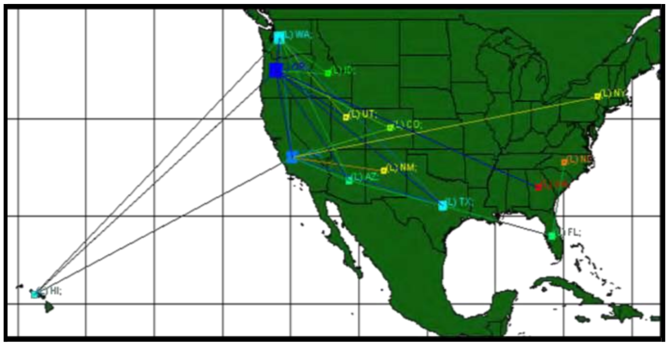

Fig. 1. Visualization of the top 20% of the network – data filtered to remove edge weights < 37.

51. The average advertised age was 25 (SD = 5.83). Of the 603 unique providers, 236 advertised their ethnicity. The top advertised ethnic groups included: Asian (30%), Hispanic/Latin (25%), two or more ethnicities (18%), and Black (11%).

IV. Findings

Of the 517 unique phone numbers observed, 346 phone numbers were advertised in multiple locations (67% of the sample). Based on the advertisement history, 47 states and 285 cities were represented. The number of advertised locations ranged from 1 to 35. The average number of advertised locations was 7 (SD = 3.00). This information was used to create a provider-by-location matrix, which allowed the movement trends to be observed using SNA.

The average weighted degree is important for this network because the edge weight describes the number of providers moving between locations (flow of movement).

TABLE I. DESCRIPTIVE STATISTICS FOR THE LOCATION NETWORK.

Measure	State Level	City Level
Nodes (n)	51	285
Edges (m)	655	4969
Average Degree	25.69	34.87
Avg. Weighted Degree	231.18	56.06
Network Diameter	5	6
Density (*d*)	.514	.123
Avg. Clustering Coefficient	.764	.756
Avg. Path Length	2.33	2.31

The locations with a higher weighted degree represent hub locations for the activity being studied. The density for the state level network was .514, which indicates 51% of all potential edges were present in the network making this a dense network. Most real-world networks are sparse. The city level density was much lower at .123, which is typical as networks grow (more nodes introduced). The average path length for the state and city networks was 2.33 and 2.31 respectively. Short average path lengths are indicative of small world properties, meaning any node in the network is able to reach any other node in a short amount of hops. Table1 below provides the descriptive statistics for each.

A. Centrality Measures

A variety of centrality measures were calculated to identify the most central or prominent nodes in the network. Identifying central nodes in the network can expose critical or vulnerable points that can be targeted to disrupt network activity. The centrality measures of main concern to this study were degree, eigenvector, and betweenness. The top 10 ranked states based on these centrality measures are presented in Table 2. A city level analysis was also run to identify hub locations within a state.

TABLE II. LOCATION NETWORK CENTRALITY MEASURES FOR THE TOP 10 STATES.

Rank	Degree Centrality	Value	Eigenvector Centrality	Value	Betweenness Centrality	Value
1	OR	.073	OR	.805	NV	.073
2	CA	.061	WA	.733	AK	.062
3	WA	.057	CA	.619	PA	.062
4	TX	.033	TX	.350	AR	.060
5	HI	.024	AZ	.254	CT	.059
6	AZ	.020	HI	.238	LA	.055
7	FL	.020	ID	.192	KS	.050
8	ID	.015	UT	.163	MI	.050
9	CO	.013	CO	.136	GA	.041
10	UT	.013	NM	.118	FL	.037

Degree centrality is useful in identifying hub locations. These nodes have the highest number of edges connecting them with other nodes (i.e. they have many neighbors). Eigenvector centrality takes into consideration a node's degree as well as the degree of its neighbors. It can be large either because a node is highly connected or it is connected to other highly connected nodes. In terms of this study, these locations had both high degree measures and were connected to each other to form a prominent circuit.

Betweenness centrality measures nodes that lie on the shortest paths between many other nodes. This implies these states potentially fall along many circuits in the network. These nodes may not at first-glance appear to be important because they may not have a high degree in relation to other nodes in the network. However, they do present vulnerable points for potential intervention.

B. Visualization

After initial analysis, the network was visualized at the state and city/county level. Thresholds were established to filter the data based on edge weight, to reduce noise and to observe the most prominent circuits. Filtering the data allowed the locations with the highest volume of providers being advertised between them to be displayed. As the thresholds were relaxed, greater portions of the network were exposed. Figure 1 depicts the movement trends at the state level.

The top 10% of movement in the network (n=5, m=5) at the state level, the prominent circuit consisted of California, Hawai‘i, Oregon, Texas, and Washington. These five hub states represent four of the seven states identified in the Western Circuit (CA, HI, OR, and WA). When thresholds were reduced to include the top 15% of the network (n=7, m=11), the circuit expanded to include Arizona and Idaho with edges becoming visible between hub nodes exposing larger portions of the Western Circuit. When thresholds were reduced to include the top 20% of the network (n=14, m=24), the circuit expanded to include Colorado, Florida, Georgia, North Carolina, New Mexico, New York, and Utah as well as additional edges between the existing nodes. At this data filter, the circuit included the entire Western circuit as well as bi-coastal movement. At the city level, thresholds were put in place representing circuits with an edge weight of 10 or greater. When data was filtered at this level, three prominent circuits were apparent representing the intra-state or micro-circuits for the advertisement source locations. As thresholds were reduced the expansion of micro-circuits to cities in neighboring states were observed.

V. DISCUSSION

This study demonstrated that content available in online escort advertisements can be used to observe movement patterns across the U.S., and that tracking phone numbers can be an effective method for detecting movement. Movement trends were analyzed using social network analysis methods to identify covert network activity, and by incorporating and visualizing geospatial data, the most prominent hub locations (state and city level) were identifiable based on the advertisement activity of the market. This method reveals a potential vulnerability of covert networks that can be exploited to observe patterns of activity.

The movement trends observed were consistent with known trafficking hubs and circuits based on retrospective law enforcement data. In contrast with retrospective data, this method could provide a tool for law enforcement and service providers to identify potential trafficking cases when providers are advertised prior to their arrival in a new city, and use this information for intervention and other anti-trafficking strategies. Degree and eigenvector centrality measures exposed the hub locations with the highest volume of movement between them. Betweenness centrality measures were useful in identifying the locations that lie on the shortest paths. They may not have the highest volume of activity, but appear to be common stops along circuits. Nodes with the highest betweenness centrality present a vulnerable point in the network and would disrupt network activity if targeted as they lie on the most paths between other nodes. The locations identified by this measure may not have been known as prominent locations based on available law enforcement statistics.

Currently, law enforcement data on sex trafficking provides retrospective information about criminal activity based on the number of arrests made. The above method outlines a way to capture movement trends of potential trafficked persons prior to criminal action, allowing for a more proactive approach to law enforcement intervention. Identification of highly weighted routes could facilitate the allocation of law enforcement resources in general. Presenting the data geospatially makes the data more useful to practitioners and policy makers by making the trends more apparent. Increased understanding of this covert network can lead to the development of better strategies of prevention, protection, and prosecution

VI. REFERENCES

[1] TIP. “Trafficking in Person's Report 2012”, 2012, U.S. Department of State, available online http://www.state.gov/j/tip/rls/tiprpt/2012/ index.h

[2] d. boyd, H. Casteel, M. Thakor, & R. Johnson, “Human trafficking and technology: A framework for understanding the role of technology in the commercial exploitation of children in the U.S.”, 2011, Microsoft Research, available online http://research.microsoft.com/enus/ collaboration/ focus/education/htframework2011.pdf.

[3] M. Latonero, "Human trafficking online: The role of social networking sites and online classifieds", 2011, available at SSRN 2045851, 2011.

[4] M. Ibanez, “Virtual red light districts: Detecting covert networks and sex trafficking circuits in the US,” 2015, (Doctoral dissertation, UNIVERSITY OF HAWAI'I AT MANOA).

[5] D.M. Hughes, “The internet and sex industries: Partners in global sexual exploitation”, Technology and Society Magazine, IEEE, 19(1), 2000, pp. 35-42.

[6] M. Latonero, “Technology and human trafficking: The rise of mobile and the diffusion of technology-facilitated trafficking”, 2012, available at SSRN 2177556, 2012.

[7] Operation Broken Silence. “The Nashville Backpage Report: An Analysis of the Online Commercial Sex Industry and Human Trafficking in Tennessee”, 2012, Operation Broken Silence, available online http://www.operationbrokensilence.org/wp-content/uploads/ 2012/01/Nashville-Backpage -Report.pdf.

[8] S. Kreyling, C. West, & J. Olson. “Technology and research requirements for combating human trafficking: Enhancing communication, analysis, reporting, and information sharing” In Washington: Pacific Northwest Laboratory, 2011, available online http://www.pnl.gov/main/publications/external/technical_reports/PNN L- 20258.pdf

[9] M. Baccara & H. Bar-Isaac, “How to organize crime” The Review of Economic Studies, 2008, pp. 1039-1067.

[10] K. Kalm, “Illicit network structures in cyberspace”, In Cyber Conflict (CyCon), 2013 5th International Conference on (pp. 1-13). IEEE.

[11] C. Morselli, C. Giguère, & K. Petit, “The efficiency/security trade-off in criminal networks”, Social Networks, 2007, pp. 143-153.

[12] M. Sparrow, “The application of network analysis to criminal intelligence: An assessment of the prospects”, Social Networks, 1991, pp. 251-274.

[13] S. Chainey, L. Tompson, & S. Uhlig, “The utility of hotspot mapping for predicting spatial patterns of crime,” Security Journal, vol. 21(1-2), pp. 4-28, 2008.

[14] L. Smith, S. Vardaman, & M. Snow, “The national report on domestic minor sex trafficking: America's prostituted children”, 2009, Shared Hope International, available online http://sharedhope.org/ resources/ TheNationalReport.aspx