

The fight against sex trafficking

Monitoring online ads, social media and other internet activities can help identify and rescue victims of this murky, complex crime. **Nickolas K. Freeman, Gregory J. Bott, Burcu B. Keskin** and **James J. Cochran** describe how their advanced analytical techniques are helping police

Human trafficking is a form of modern-day slavery that affects millions of individuals worldwide. It can take various forms, including labour trafficking, sex trafficking, and forced marriages. Farrell *et al.*¹ define human trafficking as a serious and criminal violation of human rights that involves the recruitment, transportation, harbouring or receipt of people through force, coercion, or deception for exploitation. Sex trafficking involves the use of force, fraud, or coercion for the purpose of compelling to engage in prostitution, pornography, or other forms of commercial sexual exploitation. In the United States, the definition does not depend on fraud, force, or coercion when the victim is a minor (bit.ly/49oxuQq).

Various organisations and researchers have attempted to estimate the scale of sex trafficking based on available data and anecdotal evidence. According to the United Nations Office on Drugs and Crime's *Global Report on Trafficking in Persons*,² sexual exploitation is one of the most common purposes of human trafficking, and women and girls make up the majority of victims. However, the report acknowledges that the true extent of sex trafficking is likely severely underreported. Accurately quantifying sex trafficking is challenging due to factors such as the hidden nature of the crime, the reluctance of victims to come forward, and the involvement of organised criminal networks. Furthermore, sex trafficking patterns can change over time due to various factors, including law enforcement efforts, economic conditions, and shifts in demand.

The challenges associated with estimating the global scale of sex trafficking also apply to the task of estimating its scale in

the US. The National Human Trafficking Hotline (humantraffickinghotline.org) is a valuable resource in the US for reporting and gathering information on human trafficking cases. According to its annual reports, thousands of potential cases of both sex trafficking and labour trafficking are reported to the hotline each year. The Polaris Project (polarisproject.org), a non-profit non-governmental organisation that works to prevent sex and labour trafficking in North America, shows that sex trafficking is a substantial issue in the US, with cases reported in all 50 states. However, sex trafficking prevalence varies substantially by state and locality. Some states have higher reported cases of sex trafficking due to factors such as proximity to international borders, their status as major transportation hubs, and relatively high demand.

Sex trafficking in the US, as in other parts of the world, can take various forms and is conducted through a range of methods.³ The specific tactics used by traffickers may vary, but the internet is a common platform used to facilitate sex trafficking in the US. Specifically, traffickers may use online classified ad websites, social media, and other online platforms to recruit and market victims for commercial sexual exploitation.^{4,5} The internet allows traffickers to reach broad audiences while maintaining anonymity and disguise.

Many people erroneously assume that internet ads marketing sex trafficking victims are hidden on the dark web. However, because a substantial proportion of commercial consumers lack the technical savvy needed to navigate the dark web, online advertisers often advertise on the surface web. Online platforms also often facilitate payment processing, making

it easier for customers to pay for sexual services. Payment via online platforms, gift cards, cryptocurrencies, and other methods makes it difficult to trace these financial transactions. Since the closure of the popular classified ad website backpage.com in 2018, the online ecosystem of sites hosting ads has fragmented, with many sites now operating actively in the ecosystem. Thus, organisations that previously tracked ad data from a single site now face the challenge of linking data across many sites, a daunting task given the high volume of ads posted daily (e.g., more than 100,000 ads and more than 500,000 photos per day)⁶ and the use of obfuscated language and symbols (such as emojis) within the ads.

The Institute of Data and Analytics (IDA) is housed at The University of Alabama and is composed of experts in the areas of analytics, supply chain networks, information systems, statistics, and cybersecurity. We lead an IDA initiative named STANDD (Sex Trafficking Analytics for Network Detection and Disruption: ida.culverhouse.ua.edu/initiatives/human-trafficking/), which since May 2019 has seen us collaborate with local, state, and national law enforcement organisations (LEOs) and non-profit organisations (NPOs) to develop advanced analytical techniques for detecting and disrupting sex trafficking activities using ads posted on the internet. We have developed an online classified ad data collection and analysis tool that generates data products that inform operations conducted by LEO and NPO partners in the fight against sex trafficking. The data collection component of the developed tool spans more than ten active websites in the online commercial sex ecosystem, resulting in a research data set with details on the text and

images associated with more than 200 million ads. This data is processed by an analytical pipeline that links, classifies, and filters groups of ads based on spatiotemporal advertisement patterns, text and emoji usage, and images. Instead of relying solely on text attributes (e.g., phone numbers and ad headings) for linking individual posts, the pipeline utilises state-of-the-art techniques to incorporate images into the linking process. The analytical pipeline leverages emerging tools for graph data management and analysis to perform sophisticated anomaly detection and achieve scalability (i.e., to continue to work efficiently as the amount of data under consideration grows rapidly). Data products produced by the analytical pipeline allow LEOs and NPOs to focus their limited resources on individuals who are at the highest risk for trafficking in their area.

Since 2021, we have assisted law enforcement with the arrest of more than 200 “johns” and more than 30 sex offenders for “traveling to meet a minor” as well as the identification of more than 50 victims.

Here, we explain the nature of classified ads, data collection and analysis challenges, and details of our ad-linking algorithm. Technical details of the research can be found in papers by Keskin *et al.*⁶ and Freeman *et al.*⁷.

Attributes of classified online ads

Online classified ads facilitating sex trafficking contain information that can provide insights to law enforcement, researchers, and advocacy groups. While these ads are designed to be discreet, they include

important details, including the target location (target city/target state), contact information (phone numbers, usernames, etc.), specific services (using specific language with emojis) and photographs.

Online classified ads are organised by state and city (or region) combination, as shown in Figure 1(a). Selecting a particular state and city (or region) combination opens the ad board containing several posts, as shown in Figure 1(b). Once clicked, each post link reveals additional information, including a scam warning, post date and time, contact information, details regarding services offered, price of the services, and several photographs.

Ad linking

Integrating and analysing data from multiple sources is a necessary step for building a comprehensive picture of trafficking operations. However, there are several challenges to linking ads from multiple websites. First, the cities may be different or may have different spellings (e.g., shorthand representation) on each site. Second, some location boards do not have a city name but a region encompassing several cities. These regions sometimes indicate ambiguous locations or overlapping locations. For instance, in Figure 1(a), in the state of Kentucky, there is a region, “Eastern Kentucky”, but there is also a city, Lexington, that is located in Eastern Kentucky. These variations happen even on the same website, but they are more common when collecting information from multiple websites. Finally, posted photographs of the individuals

involved can be displayed differently (e.g., in black and white or in colour), cropped, modified, or obscured across sites. Even if the same individual and the same photographs are posted to multiple websites, these modifications make it more difficult to link the same individuals. The STANDD image-processing techniques used in the analytical pipeline compare images from different sites based on perceptual hash analysis. The data is used to populate both relational and graph databases.

Anomalous linkages are another major issue that can arise from the use of generic images (site logos), generic texts (“Available Now”) and scam ads. The detection and removal of these anomalous links and data are essential for generating meaningful data products. Regarding scam posts, scammers post fake ads hoping to entice susceptible “johns” to make deposits without any delivery of services. To entice johns, scammers may use photographs they have seen in ads on other sites or platforms such as Instagram. The volume of a scam post is typically much higher than that of a post from a real provider because the scammers will often post identical ads targeting a large number of locations numerous times a day. Aside from scams, real sexual service providers may want to avoid using their own photographs and “borrow” photographs of people with similar characteristics instead (e.g., a similar build, age, and hair colour). Regardless of the cause, the presence of these data points is extremely problematic for linking ads accurately.

(a)

(b)

Figure 1: Attributes of online classified ads: (a) location boards; (b) ad boards for a specific state/city or region.



Dr. Nickolas K. Freeman is an associate professor of operations management at The University of Alabama and serves as an assistant director of the Institute of Data and Analytics.



Gregory J. Bott is the Marillyn A. Hewson Endowed Professor of Cybersecurity at the University of Alabama's Culverhouse College of Business and assistant director of the Institute of Data and Analytics.

► Ultimately, it is virtually impossible to validate the authenticity of any post or image. Thus, our approach to handling anomalous data points is to determine the distribution of image and post-text connectivity in a graph representation of our data and filter out data points that exhibit abnormal connection patterns.⁸ For this purpose, we employ a widely used graph metric called betweenness centrality (BC) as our measure of connectivity. The BC value for a node in our graph representation of the data (e.g., a specific image, post text, or contact number) captures the proportion of the shortest paths between all node pairs in the graph that must go through the target node, with higher values indicating higher connectivity. In a previous study,⁷ we show how this metric, along with the degree of nodes in the graph representation, can be used to determine thresholds for excluding data points that are likely to be erroneous. This insight allows us to filter the graph representation, removing many anomalous nodes. We then extract connected

components of the graph (i.e., groups of nodes that have some path between them), which represent data points that are likely associated with an individual or unique posting entity. An example of the graph representation for one such component is given in Figure 2, where the different colour nodes represent different data types – for example, image (orange nodes), post text (blue nodes), or contact information (green nodes).

Each graph component includes a collection of data points that correspond to the post texts, images, and contact information used by an individual or posting entity. We then map these artefacts back to the actual ads to determine when they were posted and the locations the ads targeted. This mapping allows us to investigate the ad posting patterns over space and time (i.e., spatiotemporal patterns). We use these spatiotemporal patterns to further filter the graph by eliminating components that exhibit patterns that would be impossible for an actual individual.

The overall framework, given in Figure 3, details all of the analytical steps, from data collection to the construction of data products (see the paper by Freeman *et al.*⁷ for the technical details of each step).

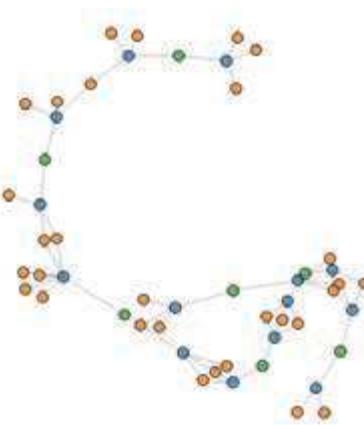


Figure 2: Connected component example.

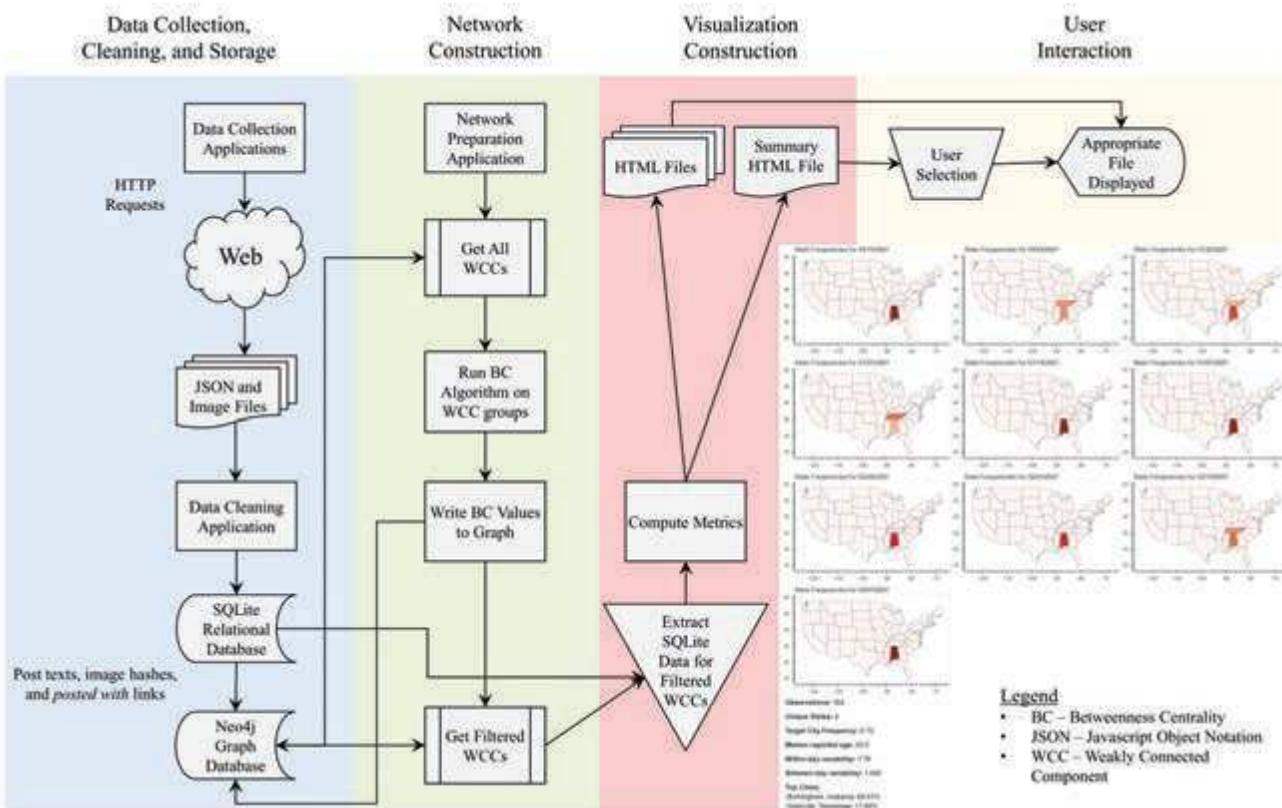


Figure 3. Overall analytical framework.



Dr. Burcu B. Keskin is a professor of operations management and the Reese Phifer Fellow at the University of Alabama, Culverhouse College of Business.



James J. Cochran is professor of statistics and Mike & Kathy Mourin Research Chair with the University of Alabama's Culverhouse College. He is a Fellow of the ASA and INFORMS.

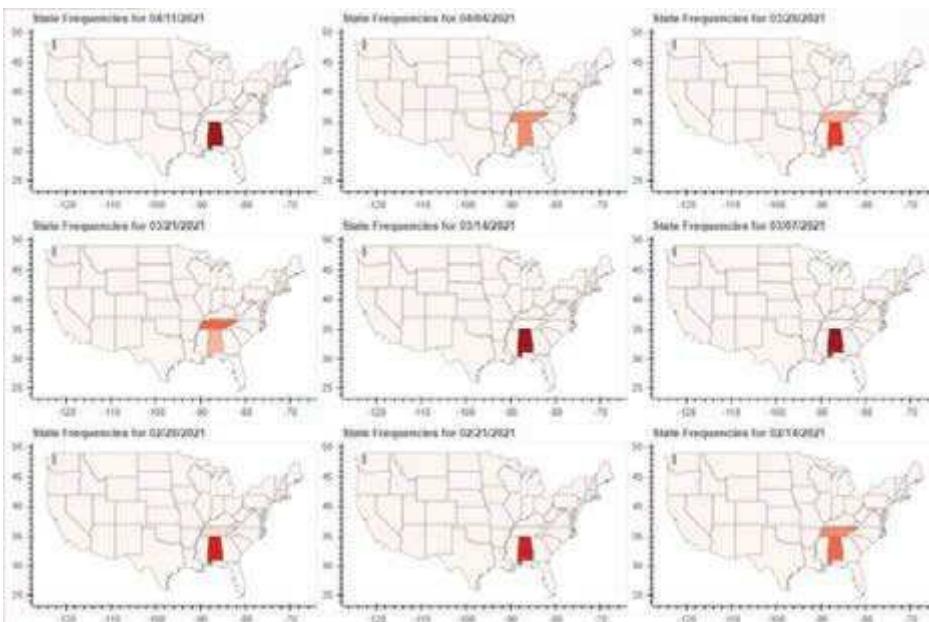


Figure 4: Spatiotemporal pattern of a suspected trafficking activity.

Putting data into action

We maintain the data collection and analysis process described and regularly provide data products for LEO and NPO partners. The insights offered by our products have been used in more than a dozen operations and assisted in identifying several potential sex trafficking victims. When planning for an operation, we monitor recent activity to identify ad groups which represent individuals exhibiting signs that are indicative of trafficking. We create spatiotemporal patterns for each group, as shown in Figure 4. Note that this figure represents 9 weeks of interstate movement, with the first week on the bottom right-hand side and the last week on the top left-hand side. The shaded states represent the posting locations for a single individual. These spatiotemporal patterns help predict the movement of individuals and provide insight on patterns that may correlate with sex trafficking (e.g., interstate movement).

Impact and concluding thoughts

The crime of sex trafficking is too hidden, complicated, and ever-changing to be tackled by one group alone.

The work done by the STANDD team demonstrates how techniques from the fields of information systems, image processing, graph theory, and pattern

recognition can transform online sex ads into data products that are able to help identify and rescue sex trafficking victims. Our LEO and NPO partners use the insights gained from data products generated by our analytical pipeline to plan and execute operations aimed at disrupting trafficking networks. We actively work with more than a dozen law enforcement and non-profit agencies in states including Alabama, Mississippi, New York, Texas, Louisiana, and Colorado. Our work involves undercover operations, surveillance, and outreach. This research also allows our partners to be more effective with limited resources.

Identified victims are connected with service providers that have the resources to assist with rehabilitation. Rehabilitating sex trafficking victims is a complex process that involves physical, psychological, and social support. It is a long and challenging process that often requires several organisations working together on a comprehensive and holistic solution to help survivors regain their lives and autonomy. Although the rehabilitation step is beyond the scope of the STANDD initiative, its partner network is connected to various organisations, government agencies, and non-profits that are able to tailor rehabilitation efforts to the specific needs of each survivor.

We actively work with more than a dozen law enforcement and non-profit agencies. Our work involves undercover operations, surveillance, and outreach

Based on our experience, addressing the problem of sex trafficking requires a multi-faceted approach that encompasses policy, law, technology, public awareness, and victim support. Law enforcement, government agencies, academic scholars, technology companies, non-profit organisations, and the public must work together to disrupt sex traffickers and help survivors escape, recover and rebuild their lives. ■

References

1. Farrell, A., DeLateur, M. J., Owens, C., and Fahy, S. (2016) The prosecution of state-level human trafficking cases in the United States. *Anti-Trafficking Review*, (6), 48–70.
2. United Nations Office on Drugs and Crime (2022) *Global Report on Trafficking in Persons 2022*. United Nations. bit.ly/472H6ic
3. Baird, K. and Connolly, J. (2023) Recruitment and entrapment pathways of minors into sex trafficking in Canada and the United States: A systematic review. *Trauma, Violence & Abuse*, **24**(1), 189–202.
4. Whitcomb, D. (2019) Exclusive: Report gives a glimpse into the murky world of U.S. prostitution in the post-Backpage era. *Reuters*, 11 April. bit.ly/3QphncR
5. Gezinski, L. and Gonzalez-Pons, K. M. (2022) Sex trafficking and technology: A systematic review of recruitment and exploitation. *Journal of Human Trafficking*. dx.doi.org/10.1080/23322705.2022.2034378
6. Keskin, B. B., Bott, G. J. and Freeman, N. K. (2021) Cracking sex trafficking: Data analysis, pattern recognition, and path prediction. *Production and Operations Management*, **30**(4), 1110–1135.
7. Freeman, N. K., Keskin, B. B. and Bott, G. J. (2022) Collaborating with local and federal law enforcement for disrupting sex trafficking networks. *INFORMS Journal of Applied Analytics*, **52**, 395–470.
8. Janson, S., Knuth, D. E., Łuczak, T. and Pittel, B. (1993) The birth of the giant component. *Random Structures and Algorithms*, **4**(3), 233–358.