Detection of Illicit Massage Businesses Through Spatial and Socio-Demographic Data Enrichment

LEAD PI: MARIA MAYORGA

Criminal Investigative Processes

SUMMARY

Illicit massage businesses, which are often associated with human trafficking activities, can be difficult to detect because they use the same online platforms as legitimate massage businesses, such as advertisement services, job recruitment ads, and review boards. Investigators need better tools to help identify illicit massage businesses to combat human trafficking and public health risks. This project will use common online review sites to develop data analysis techniques that enable automatic detection of illicit massage businesses, and the tools and techniques developed in this project may subsequently be applied to other illicit business detection as well.

PROBLEM STATEMENT

Over 10,000 illicit massage businesses (IMBs) are scattered throughout the United States, many of which force women into sex trafficking. Current efforts to identify IMBs heavily rely on massage-specific sex buyer forums. However, online evidence of trafficking risk factors can also be gathered from more static review sites such as Yelp and Google. Customer reviews and business information can be mapped to other data sources such as business licenses and court records. By mining for correlations between labeled massage business truth sets and external data including zoning, census data, business hours, similar authorship, and more, the researchers will be able to pivot between known IMBs and identify new potential IMBs. This project focuses on the state of Colorado which has an estimated 200 IMBs.

APPROACH

The researchers create classification models for detecting risk factors of human trafficking in massage business internet data using methods such as natural language processing and machine learning. The team relies on statistical methods and domain knowledge to engineer and select features from several data sources. Preliminary analysis using these methods was conducted in other states. As the work expands to Colorado, the team will use an active learning technique to adapt existing methodology to new data without having to manually create an extensive ground truth training data set.

The team will develop models that can effectively understand the language and business patterns that are unique to IMBs in Colorado. By testing this framework in Colorado, the researchers can eventually expand this analysis nationwide without sacrificing performance.

RESULTS

The researchers identified around 2,600 massage businesses in Colorado with reviews on Yelp and around 200 on Rubmaps. They have further identified over 100 businesses with potential trafficking risk factors. For example, the reviews contain terms related to trafficking or the business's phone number is used in commercial sex ads. The team will examine these businesses and assign labels to them as a starting point. They have also located records for 25,000 Colorado massage therapists and are working to identify which ones have received disciplinary actions related to human trafficking as well as their practice locations. Next, the team will use the Rubmaps data and de-identified therapist records to create a ground truth data set of the most likely IMBs and begin to build risk prediction models (e.g. decision trees) for IMBs in Colorado.

ANTICIPATED IMPACT FOR DHS

This project aligns with multiple DHS priorities related to homeland security; cyberspace and infrastructure security; economic security; border security; and countering human trafficking and exploitation. Illicit businesses tend to operate within organized criminal networks. In this work, the team may uncover connections between IMBs through online networks linked to transnational organizations. The transnational nature of trafficking requires swift coordination and data sharing within border forces and their international counterparts. Identifying potential IMBs and exposing criminal use of infrastructure (e.g., truck stops) can lead to the removal of victims from danger. In addition to the human toll, the economic impact of IMBs is large, with estimates of total annual revenue in the U.S. over $3 billion.