

TRAFFICBOARD: Digital Spatio-Temporal Pinboard for Human Trafficking Detection

Catalina Vajiac* Duen Horng Chau† Andreas Olligschlaeger‡ Pratheeksha Nair§ Meng-Chieh Lee*
Mirela Teixeira Cazzolato¶ Reihaneh Rabbany§ Christos Faloutsos* Cara Jones||

1. Expert uses **Graph view** to discover connections between metadata and select nodes of interest
2. Aligned timeseries in **Timeline view** let experts examine synchronized behavior and select particular weeks
3. Expert finds hotspots in **Map view** and can **select** a state to zoom in and investigate city-level data

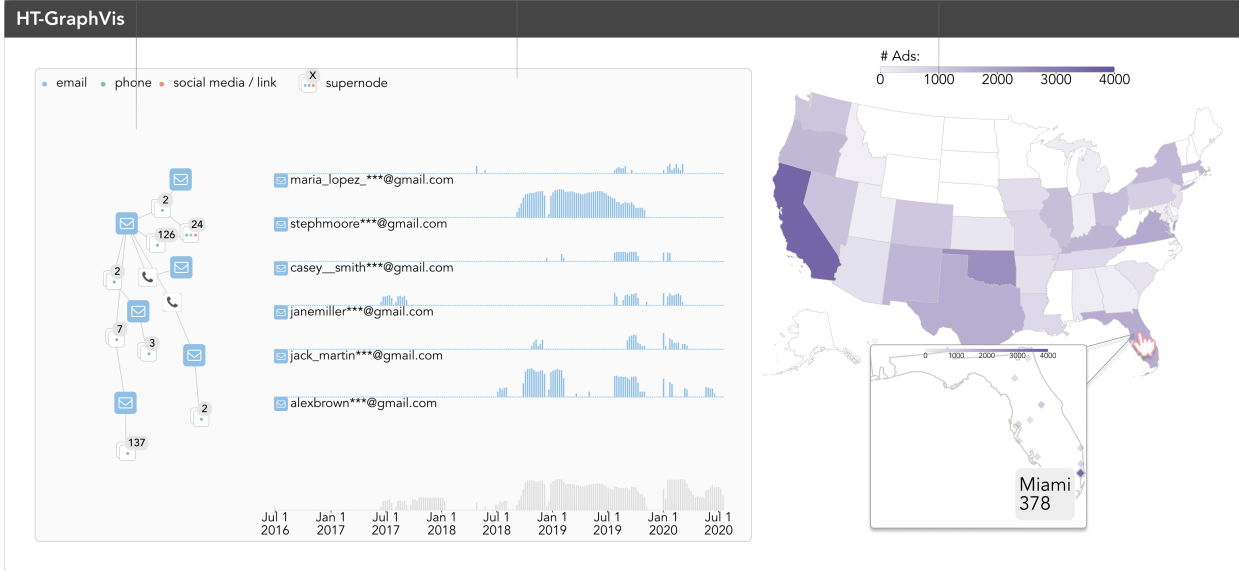


Fig. 1. **TRAFFICBOARD** showing real-world data (entity names obfuscated for privacy): **1. Graph view** lets experts quickly see connections between different metadata types. **2. Timeline view** shows time-based data of each metadata node. **3. Map view** shows geographical spread of data. All views are linked, so selections in one view updates all data automatically.

Abstract—Practitioners can detect human sex trafficking (HT) in online escort websites by analyzing both the text and connections between metadata (i.e. contact information) throughout time of suspicious clusters of connected ads. Previous attempts at visualizing these graphs have suffered from the hairball effect, as certain portions of the graph are densely connected. Given a group of ads with shared metadata, how can we interactively visualize connections between metadata and how they change over time? TRAFFICBOARD combines graph compression techniques and interactivity to enable practitioners to investigate suspicious metadata and their connections throughout time, as well as their geographic spread.

Index Terms—Human Trafficking, Spatio-Temporal, Node-Link Diagrams, Time-evolving Graphs.

1 INTRODUCTION

Human trafficking for forced sexual exploitation (HT) is incredibly pervasive, affecting an estimated 6.3 million people a year [3], and the majority of HT victims are advertised on online escort websites [5]. At-will sex workers also post on these sites, and investigators try to separate HT rings from at-will workers. There is one critical insight to detecting HT; since traffickers entirely control the ad content for their victims [5], ads posted by the same trafficker tend to be similar,

including similar phrasing and common metadata between ads.

Analysts, such as law enforcement officers, criminologists, or NGO workers, often utilize connections between metadata to analyze suspicious clusters. Previous work on HT evidence visualization [6] has stayed away from node-link diagrams, as these graphs contain very densely connected regions that make them unreadable. However, investigators are very familiar with this visualization technique. In particular, investigators commonly use an analog version of node-link diagrams called an Anacapa chart [2, 4], a collage of media pinned to a pinboard and connected with red string to mark connections that is commonly featured in detective films. How can we digitize the Anacapa board, easily scaling it to hundreds of nodes or more, as well as including spatio-temporal information?

TRAFFICBOARD addresses the following design challenges C1 – C2 created in conjunction with domain experts:

- C1. **Summarizing the display graph.** Node-link diagrams are notorious for being harder to read as the number of nodes and edges increase, however, this is the graph visualization analysts are most

*Carnegie Mellon University: {cvajiac, mengchil, christos}@cmu.edu

†Georgia Institute of Technology: polo@gatech.edu

‡i3 LLC: olligschlaegerandreas@gmail.com

§Mila & McGill University: {pratheeeksha.nair, rrabba}@mail.mcgill.ca

¶University of Sao Paulo: mirelac@usp.br

||Marinus Analytics: cara@marinusanalytics.com

comfortable with. How can we construct a graph that will be easy to process, even if we have hundreds of nodes?

- C2. **Fast spatio-temporal exploration.** We not only have metadata connectivity, but also geographic and time data associated with each ad. How can we allow analysts to intuitively explore combinations of this data?

Reproducibility: The code is available at <https://github.com/catvaji/trafficboard-vis-2023-poster>, with a synthetic data generator so others can try TRAFFICBOARD without compromising victims' safety.

2 DESIGNING TRAFFICBOARD

Our design challenges and TRAFFICBOARD were designed through 6 months of participatory design. We discuss how we address challenges C1: *Summarizing* and C2: *Spatio-Temporal* in Sections 2.1 and 2.2.

2.1 Addressing C1: Creation of display graph

To use the node-link diagram format investigators are most familiar with, we compress the original input graph to make the size of the display graph more manageable while maintaining the overall structure of the original graph. To create the display graph from escort advertisements, we do the following steps displayed in Figure 2).

- (a) **Construct metadata graph.** We start by constructing an undirected graph $G = (V, E)$ where each $v \in V$ represents a phone number, email, or link / social media account used in the ads and each edge $(u, v) \in E$ represents at least one ad that contained both u and v in the data.
- (b) **Find central nodes.** Since these graphs often have many nodes, we want a succinct way to summarize their overall structure so the resulting graph will be simple to display. To this end, we first introduce *central nodes*, which intuitively are nodes that are structurally important to the overall input graph. We find *central nodes* C by picking the top 6 nodes in G (due to space constraints) with highest betweenness centrality [1].
- (c) **Find supernodes.** Once we've found central nodes, we then compress some nodes into *supernodes* if they are connected to the same central nodes. More specifically, let $N_C(u)$ represent the neighborhood of u where for all v in $N_C(u)$, $v \in C$. Then, for two $v, w \in V$ where $v \notin C$ and $w \notin C$, we combine them into a supernode s if $N_C(v) = N_C(w)$.

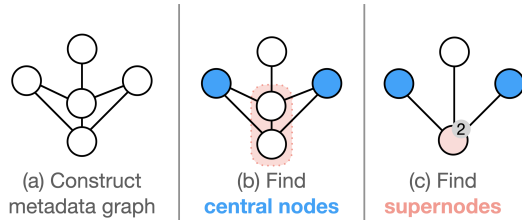


Fig. 2. *Creation of display graph:* (a) connect metadata nodes if they are both present in any ad, (b) find the *central nodes* that are most important to the graph structure, and (c) create *supernodes* by combining any nodes whose neighborhoods contain the same central nodes.

2.2 Addressing C2: Design and Usage Scenario

To address this challenge, we combine node-link diagrams with time-series and geographic data, displaying the most important timeseries by default and allowing the rest to be explored as details on demand.

Figure 1 demonstrates how an analyst (i.e. investigator, social worker, criminologist) may use TRAFFICBOARD to investigate connections between metadata. TRAFFICBOARD is comprised of three main views: *Graph view*, *Timeline view*, and *Map view*. The analyst will first investigate *Graph view*, observing six highlighted central nodes. She may notice that within *Graph view*, the display graph is shown with different icons for each node type (phone number, email, link, and supernode). Each supernode shows the number of nodes it compresses and, through the colored dots in the center, their node types. Central nodes are highlighted in the respective node type's color. The analyst may select any node to subset the data in the rest of the interface.

The analyst may then look at *Timeline view*, seeing the aligned timeseries for each central node. She may use the time selector at the bottom to select a date range, which will then highlight the nodes that were in posted advertisements during that period. Then, she might select a supernode in *Graph view*, which will then display the timeseries of each node contained in that supernode as shown in Figure 3.

When an expert **selects** a supernode...

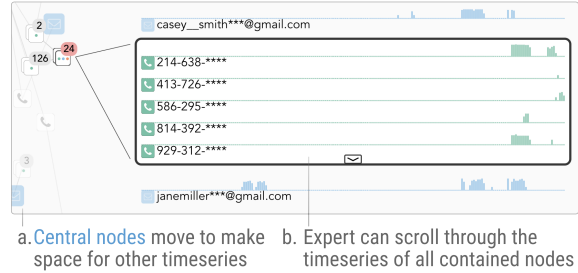


Fig. 3. **Selection of a supernode:** analysts can explore the timeseries of all nodes contained within a supernode by simply selecting it – then all data will be displayed. PII obfuscated to maintain victim privacy.

At any time during her investigation, the analyst may choose to look at *Map view*, which by default, displays a choropleth map of the number of ads in each location. If the analyst has previously selected a node in *Graph view* or a time range in *Timeline view*, the map will have updated to only show data that survive this filters. The analyst may select a state, the specific city data of that state is shown. Then, she may click on a city to subset the data in all other views.

3 CONCLUSION

TRAFFICBOARD was built not only to digitize the Anacapa charts (i.e. analog node-link diagrams) investigators are most familiar with, but also to enable them to investigate changes in connectivity over space and time. We believe TRAFFICBOARD makes two major contributions:

- C1. **Summarizing the display graph** to allow node-link diagrams be used for HT data without overwhelming the investigator; and
- C2. **Fast spatio-temporal exploration** to allow investigators to explore how the usage of different metadata changes over time and geographical location.

So far, we have collaborated with two domain experts for our design. We plan to solicit feedback from more experts, then conduct a user study with investigators to assess the efficacy of TRAFFICBOARD.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE1745016 and DGE2140739, as well as the Pittsburgh Infrastructure Technology Alliance.

REFERENCES

- [1] U. Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Soc. Networks*, 30(2):136–145, 2008.
- [2] W. R. Harper and D. H. Harris. The application of link analysis to police intelligence. *Human Factors*, 17(2):157–164, 1975.
- [3] I. L. Organization. Global estimates of modern slavery: forced labour and forced marriage. *International Labour Organization*, 2021.
- [4] M. K. Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. *Social networks*, 13(3):251–274, 1991.
- [5] Thorn. Report on the use of technology to recruit, groom and sell domestic minor sex trafficking victims. https://2715111qnwey246mkc1vzqg0-wpengine.netdna-ssl.com/wp-content/uploads/2015/02/Survivor_Survey_r5.pdf, 2015.
- [6] C. Vaji, D. H. Chau, A. M. Olligschlaeger, R. Mackenzie, P. Nair, M. Lee, Y. Li, N. Park, R. Rabbany, and C. Faloutsos. Trafficvis: Visualizing organized activity and spatio-temporal patterns for detecting and labeling human trafficking. *IEEE Trans. Vis. Comput. Graph.*, 29(1):53–62, 2023.