# Determining Relative Airport Threats from News and Social Media

**Rupinder P. Khandpur**[*]
Virginia Tech
Arlington, VA 22203

**Taoran Ji**[*]
Virginia Tech
Falls Church, VA 22043

**Yue Ning**
Virginia Tech
Arlington, VA 22203

**Liang Zhao**
George Mason University
Fairfax, VA 22030

**Chang-Tien Lu**
Virginia Tech
Falls Church, VA 22043

**Erik R. Smith**
ANSER
Homeland Security Group
Falls Church, VA 22041

**Christopher Adams**
ANSER
Homeland Security Group
Falls Church, VA 22041

**Naren Ramakrishnan**
Virginia Tech
Arlington, VA 22203

### Abstract

Airports are a prime target for terrorist organizations, drug traffickers, smugglers, and other nefarious groups. Traditional forms of security assessment are not real-time and often do not exist for each airport and port of entry. Thus, homeland security professionals must rely on measures of attractiveness of an airport as a target for attacks. We present an open source indicators approach, using news and social media, to conduct relative threat assessment, i.e., estimating if one airport is under greater threat than another. The three ingredients of our approach are a dynamic query expansion algorithm for tracking emerging threat-related chatter, news-Twitter reciprocity modeling for capturing interactions between social and traditional media, and a ranking scheme to provide an ordered assessment of airport threats. Case studies based on actual aviation incidents are presented.

## Introduction

Aviation security policy makers and practitioners continue to face a demanding security environment. Terrorism, security breaches, smuggling, and increasingly sophisticated cyber intrusions capable of disrupting essential services are just a few examples of the threats faced by airports. Unfortunately credible, consistent, and real-time reporting on threats does not exist for all homeland security assets (airports, ports of entry, and other critical infrastructures) uniformly.

So how can we provide a more comprehensive understanding of threats to better inform those decisions that guide resource allocations and security measures? We hypothesized that if we consider a broad range of open source indicators (OSI), we can account for some deficiencies in specific intelligence, as well as incorporate public opinions and concerns about security directly related to homeland security. Here we present aviation security as a case study. Using social media and news sources as surrogates for threat reporting, we can draw insights about the relative attractiveness of each asset as a threat target using data analytic tools and techniques. By combining qualitative and quantitative analytic approaches, it must be feasible to derive relative threat assessments, or an estimate of how much threat is faced by a specific asset versus another similar one (Ramo 1994). This approach is thus not intended to determine an absolute percentage of likelihood that an attack occurs, or even to forecast an attack, but can be used to gain additional situational awareness, as well as resource planning, security mobilization, and improve communication between enforcement agencies.

## Background Questions

We pose and then answer the below questions to undergird the validity of our analytic method. The answers are based on interviews with several homeland security professionals who are familiar with homeland security analytics and reviews of official reports from agencies, such as the Transportation Security Administration (TSA).

**What should be the objective of a threat assessment system?** Providing information about the degrees of threats associated with on-going events and security climate can be a useful goal independent of predicting the occurrence of an actual threat. For instance, based on historical government data (such as seizures or arrests) and other airport-specific criteria (passenger or cargo flow volumes), certain airports are logically and data-supported as primary concerns. However, once we get beyond the major airport hubs, determining the threat level faced by mid-level airports becomes less intuitive and therefore benefits from data-supported analytic approaches. This information can help government agencies plan resources accordingly.

**Should we cast this as a regression or prediction problem of airports $\mapsto$ threat scores?** Focusing on threats in isolation, rather than collectively, may result in a misdirection of resources to less serious problems. The Department of Homeland Security risk management doctrine recommends that agencies develop processes that "should facilitate the ability to compare risks, as required, across the organization." Experiences from the state of California found ranking methodologies useful for comparing environmental risks and these methods can apply to comparisons of other varieties of threats. Thus the objective is to (1) assess and rank threats (2) critique threat modeling and (3) explore alternatives for mitigation and priority-setting (Ramo 1994).

**How relevant are open source indicators (news and social media) for this undertaking? Is the objective to reproduce DHS rankings of threats using other sources?**

[*] Authors have equal contribution.

A number of agencies, including the TSA Office of Intelligence, already gather information from several social media sites to mitigate threats and to promote situational awareness. The goal of tapping online social media is to cast a wider net and to help develop a proxy indicator of public opinion across the region. Therefore, open source social media indicators can offer necessary context for policy makers in assessing the comparative threat across various critical infrastructure.

**Are automated tools necessary?** The volume and diversity of social media indicators makes manual collection and analysis daunting. Automated tools are necessary to handle these challenges.

**Why is real-time analysis important?** Analyzing real-time news and social media feeds can enhance the existing investigative process and provide DHS greater clarity and visibility into possible nefarious activity and connections by providing an additional tool set. Certain existing threat assessment processes are not set up to be real-time systems that track emerging events. Instead, they are intended to look at which airports might draw a threat actor to attack or exploit an airport's conveyances, facilities, employees, passengers, cargo, or surrounding area. Real-time news and social media feeds can provide greater granularity at the airport level.

Motivated by this assessment, we develop a strategy for relative threat identification and ranking of airports. The contributions of our work are:

1. We articulate the problem of relative threat identification and present the first automated system, to the best of our knowledge, that provides an integrated situational assessment of relative airport threats from open source indicators such as news and social media.
2. We employ a dynamic query expansion methodology with news-Twitter reciprocity modeling to track evolving and complex threat-related discussions on mutually reinforced multi-source data.
3. We develop a ranking strategy to provide an ordered assessment of airport threats and calibrated to support situational awareness and resource planning while at the same time being responsive to developing situations.
4. Through various case studies on the US domestic airport system, we highlight the ability of our system to produce rankings that reflect developing security situations, such as bomb threats, shootings, and airport diversions.

## System Overview

Our proposed system is shown in Figure 1. It can be factorized into the stages of enrichment, dynamic query expansion, news-Twitter reciprocity modeling, and developing rankings of relative airport threats. These steps are detailed next.

### Data Ingest & Enrichment

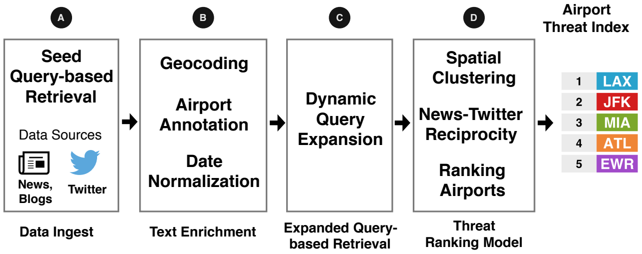


Figure 1: An overview of the relative threat ranking system.

Table 1: Seed keywords for relative threat assessment.

| Category | Seed Keywords |
|---|---|
| Threat | bomb, gun, scream, smuggle, knife, shot, threat, weapon, attack, crash, suspicious, terror, screening, scare, police, hoax, prank |
| Terrorism | dawla, baqiyah, shami, caliphate, muhajir, ghuraba, islamic state, isis, isil, daish, daash, dabiq, kuffar, kafir, takfir, sharia |
| Airport | flight, tsa, air(port\|line\|plane\|craft), terminal |
| **Final Query**: (Threat OR Terrorism) AND Airport | |

Data ingest is done using a broad class of threat-related filters to query tweets (from Topsy API) and news articles (from online web URLs mentioned in collected tweets) from January 2013 through August 2015. These filters are presented in Table 1, and were identified with help of domain experts[1]. Since the model is intended to capture more than just terrorism, these seed words serve as a starting point for broader collection of possible social media information related to airport threats.

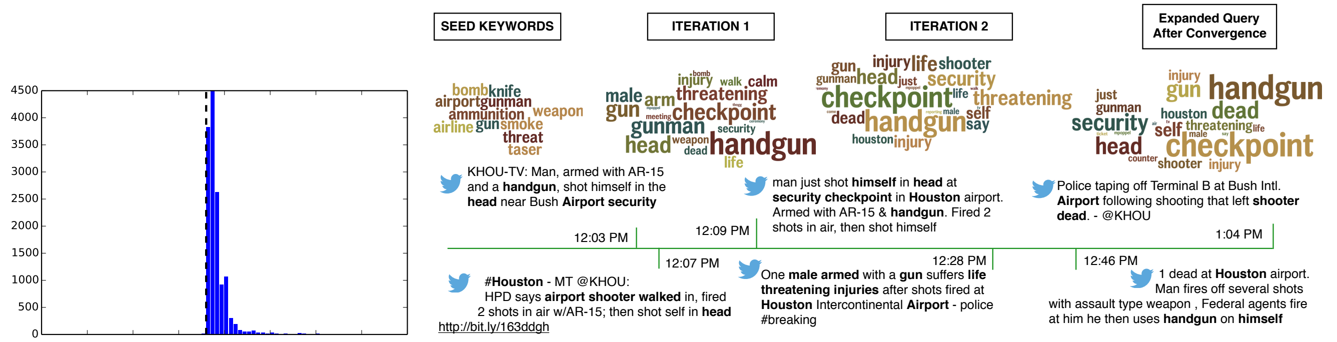
Using Basis Technology's Rosette Linguistics Platform[2] (RLP), the documents were subject to tokenization, lemmatization, and named entity extraction. Further enrichment includes geocoding, airport annotation, and date expression normalization. For geocoding, we exploit both text features and metadata of a document such as (GPS) geotag, and descriptive metatags (e.g., in the html source of news articles) to extract multiple indicators which we then use to perform spatial queries in gazetteers (constructed using GeoNames[3]) to ultimately yield the best (latitude and longitude) geolocation. We also built a large regular expression library that identifies IATA codes, canonical names, and aliases for airports mentioned in text. For instance, the SeattleTacoma International Airport (SEA) is frequently mentioned as 'SeaTac'. We use this library along with the geolocation information to map documents (tweets and news articles) to the 45 busiest U.S. airports (by total passenger boardings[4]). Finally, the TIMEN (Llorens et al. 2012) package is used to normalize date expressions.

[1] Analytic Services Inc., VA

[2] www.basistech.com/text-analytics/rosette

[3] www.geonames.org

[4] en.wikipedia.org



KHOU-TV: Man, armed with AR-15 and a **handgun**, shot himself in the **head** near Bush **Airport security**

**#Houston** - MT @KHOU: HPD says **airport shooter walked** in, fired 2 shots in air w/AR-15; then shot self in **head** http://bit.ly/163ddgh

man just shot **himself** in **head** at **security checkpoint** in **Houston** airport. Armed with AR-15 & **handgun**. Fired 2 shots in air, then shot himself

One **male armed** with a **gun** suffers **life threatening injuries** after shots fired at **Houston** Intercontinental **Airport** - police #breaking

1 dead at **Houston** airport. Man fires off several shots with assault type weapon , Federal agents fire at him he then uses **handgun** on **himself**

Police taping off Terminal B at Bush Intl. **Airport** following shooting that left **shooter dead**. - @KHOU

Figure 2: Illustration of news-Twitter reciprocity (left) and dynamic query expansion (right) around the shooting incident at Houston Airport (IAH) on 5 May 2013. We show how the news reporting (vertical dotted line) closely overlaps with bursts in Twitter activity (blue color bars) which are then dynamically tracked (see word clouds) by our model over time.

## Dynamic Tracking

Our social media data is modeled as a collection of time-ordered tweets $\mathbb{D} = \{\mathbb{D}_1, \mathbb{D}_2, \ldots, \mathbb{D}_T\}$ organized alongside $T$ time slots. Each subcollection $\mathbb{D}_i$ can be modeled as a heterogeneous graph $G = \{V, E, W\}$, where nodes $V = \mathcal{D} \cup \mathcal{F}$ consist of two kinds of nodes: tweets $\mathcal{D}$ and features $\mathcal{F}$ (tokens, hashtags, user mentions). Furthermore, $\mathcal{D}^+$ is used to denote the target Twitter subspace comprising threat-related tweets, and $\mathcal{D}^- = \mathcal{D} - \mathcal{D}^+$ denotes the unrelated set. $E$ is the set of edges formed by a tweet node and the feature nodes it contains. $W$ is the set of weights for nodes where higher weights denote a higher degree of relation between the node (either a tweet or a feature) and threat-related theme. We posit that, in targeted theme tracking, $\mathcal{D}$ and $\mathcal{F}$ are influenced by each other, which translates to tweet nodes with higher weight being more likely to have an edge with features nodes (such as bomb, gun) with higher weight. In an ideal situation, threat-related tweets could easily be collected, using higher-weighted feature nodes as the query. However, in practice, especially in our noisy and dynamic medium, a static or fixed keyword set (even if provided by domain experts) does not guarantee optimal performance for tracking evolving themes.

Leveraging this heterogeneous network $G$, we developed a tracking method based on dynamic query expansion (DQE) (Zhao et al. 2014) techniques, to search for threat-related feature nodes $\mathcal{F}^+ \subseteq \mathcal{F}$, given only a small set of seed keywords. Let $\mathcal{Q}^{(k)}$, $\mathcal{D}^{(k)}$ and $\mathcal{F}^{(k)}$ denote a set of query terms, tweet nodes, and feature nodes at the $k$th iteration, respectively. During the initialization procedure (lines $1-2$ in Algorithm 1), the theme-related query $\mathcal{Q}^{(0)}$ is initialized with $S$, its weight $W(\mathcal{Q}^{(0)})$ as a ones vector, and $\mathcal{D}^+$ as a set of tweets that match $\mathcal{Q}^{(0)}$. For the $k$th iteration, where $k \geq 1$, as shown in lines $4-6$ in Algorithm 1, the new expanded query is set as $\mathcal{Q}^{(k)}$ by selecting the *top N* weighted entities in $\mathcal{F}^{(k)}$:

$$W(\mathcal{F}) = B \cdot C \cdot W(\mathcal{D}), \tag{1}$$

where $B$ denotes the inverse document frequency (IDF) matrix of $\mathcal{F}$, and $C$ is the adjacency matrix between $\mathcal{F}$ and $\mathcal{D}$. Note that this is actually a variant of the popular TFIDF term weighting strategy. Then, as shown in lines $7-10$ in Algorithm 1, based on the weights of feature nodes, DQE selects a target subspace $\mathcal{D}^+$ such that every tweet node in $\mathcal{D}^+$ has a higher score than the tweet node in $\mathcal{D}^-$, based on the following weighting strategy:

**Algorithm 1:** Dynamic Query Expansion Algorithm

```
   Input: Seed query $S$, Twitter subcollection $\mathcal{D}$
   Output: Expanded Query $\mathcal{Q}$
1  Set $\mathcal{Q}^{(0)} = S, W(\mathcal{Q}^{(0)}) = \vec{1}$;
2  Set $D^+ = match(\mathcal{Q}^{(0)}, \mathcal{D}), k = 0$;
3  repeat
4      $k = k + 1$;
5      $W(\mathcal{F}^{(k)}) = B \cdot C \cdot W(\mathcal{D}^{(k-1)})$;
6      $\mathcal{Q}^{(k)} = topn\big(W(\mathcal{F}^{(k)})\big)$;
7      repeat
8          $swap\Big(\min\big(W(D^+)\big), \max\big(W(\mathcal{D}^-)\big)\Big)$;
9          $\sigma = \min\big(W(D^+)\big) - \max\big(W(\mathcal{D}^-)\big)$;
10     until $\sigma \geq 0$// adjust subspace;
11 until $W(F_t^{(k-1)}) = W(F_t^{(k)})$// DQE iteration;
12 $\mathcal{Q} = \mathcal{Q}^{(k)}$;
```

$$W(\mathcal{D}) = \Phi \cdot C' \cdot W(\mathcal{Q}), \tag{2}$$

Here, $C'$ is the transpose matrix of $C$, and $\Phi$ is the normalized coefficient that makes sure the weights are normalized. In this way, DQE will iteratively expand the query $\mathcal{Q} = \mathcal{F}^+$, as shown in Algorithm 1. The number of iterations required on average is very small.

## Threat Ranking Model

Our threat ranking approach is based on the joint modeling of news-Twitter reciprocity and the tweets collected from DQE. Prior to threat scoring, we first aggregate target tweet subspaces spatio-temporally, denoted by $\mathcal{D}^+_{t,a}$ where $a \in A$ refers to a specific airport which is used for spatial clustering (Ester et al. 1996). We adopt the following clustering criterion:

$$\mathcal{D}_{i,a} = \bigcup_{\lfloor \frac{t}{\varsigma} \rfloor = i} \mathcal{D}_{t,a}, \quad (3)$$

where the parameter $\varsigma$ controls window size of temporally aggregation of $\mathcal{D}_{i,a}$. In our methodology, we begin with daily, spatially clustered tweets and $\varsigma$ is set to 7 which leads to weekly aggregates.

**News-Twitter Reciprocity Modeling**: One straightforward approach to rank airport threat levels is by their cluster size, but this strategy may lead to bias due to several factors. For instance, tweets (within each cluster) may have a higher degree of relatedness to threats; passenger traffic at airports is typically varied and can influence the size inferred target subspace; and, finally, the noisy nature of Twitter can lead to inaccurate threat scoring. We overcome this by applying reciprocity modeling that consolidates the information shared across the mainstream news wire and social media, which in turn helps us standardize reporting levels.

To model this interaction, we first generate a keyword list for a given news article by combining the top 10 keywords from the tokenized content of tweets that mention a URL (pertaining to news) and the named entities mentioned in the article. Then, a TFIDF based filtering method (Ning et al. 2015) is used to extract a set of tweets associated with this news, posted within a time period. We apply this procedure for every news article, to build a set of (reciprocal) tweets $H$. The alert value of a tweet can then be determined by this interaction model:

$$v(d) = \begin{cases} \alpha \cdot W(d), & \text{if } d \in \mathcal{D}_{i,\cdot} \cap H, \\ (1-\alpha) \cdot W(d), & \text{else.} \end{cases} \quad (4)$$

where $v(d)$ denotes the alert value of tweet $d$, $\alpha$ is the coefficient used to trade off between influence of unmapped tweets and the mapped news-related tweets. Higher $\alpha$ will lead to news-related tweets making a larger contribution to the airport threat ranking.

**Ranking Airports**: By considering both the alert value of tweets for the current time window and the historical assessments, our approach to track the airport threat level on $i$th time window can be formulated as:

$$K(\tau, a) = \gamma K(\tau - 1, a) + (1-\gamma) \sum_{d \in \mathcal{D}_{i,a}} v(d), \quad (5)$$

where $K(\tau, a)$ is the threat level of airport $a$ at time $\tau$, $\gamma$ determines the weight for the current normalized threat value at this airport. Then a ranking for each time window is generated according to this threat value.

# Evaluation

Our evaluation focuses on the following key questions:

1. Do the airport rankings inferred by our approach correspond with known ground truth? How does our system's performance fare against a baseline approach?
2. Is the automated system consistent in continuously monitoring threat incidents for each airport, using multiple data streams? That is, does it produce relevant measurements of threat at each airport, over time?
3. How easily can we uncover emergent threats from the airport threats ranking?

## Evaluation Setup

The data from January 2013 through December 2013 was used as a training set for tuning parameters in our proposed open source indicators (OSI) based airport threats model: $\alpha$ and $\gamma$ (described in Equation 4 and 5). We learn two sets of parameter settings; in the first setting (P1) where $\alpha = 0.1$ and $\gamma = 0$, the influence of news reciprocity and historical priors on the threat scoring of airports is minimal. In the second setting (P2), we set $\alpha = 0.5$ and $\gamma = 0.2$, which allows higher influence of both attributes.

**Gold Standard Report:** In the absence of any official ground truth or rankings to airports threat levels we developed our own gold standard report (GSR). It contains detailed records of known airport (threat-related) incidents that took place in the United States, between 2013 and 2015. This GSR was prepared using an independent (online) source[5] and manual web search. Each GSR event contains the date of incident, airport (IATA code), description, url of the news article referring the incident, and a (manually annotated) threat category. An example of a labeled GSR event is given by the tuple (DATE="2013-11-01", AIRPORT="LAX", DESCRIPTION="TSA Agent Killed - Gunman In Custody", URL="tinyurl.com/laa6edq", TYPE="Shooting" ). We filtered this GSR for those 45 airports that were considered in our study, yielding a total 198 GSR events, as shown in Figure 3. Those events which described ongoing investigations or follow up reports to known incidents were excluded. Also, 12 airports had no associated GSR events, and were excluded from our evaluation.

## Metrics and Ranking Baselines

We make use of three metrics to measure the performance of ranking approach. They are aimed at capturing relevancy and coverage of airport rankings generated from our system.

**Average Rank (M1):** If an airport appears in a daily ranked list we assume an incident has been detected for that airport. We then construct a maximum weighted bipartite matching between the set of detected incidents ($I$) and GSR events ($E$), where allowable edges are those that satisfy an inclusion criteria, and weights on these edges denoted by their quality scores. The quality score (QS) of a match is simply the sum of a location score and a date score. Location score (LS) is defined using in a straightforward manner: if the airport names of an $(I, E)$ pair match, then LS has a value of 1, else it is 0. Date score (DS) is defined as:

$$\text{DS} = 1 - (\min(\text{date offset}, \text{window}) / \text{window}).$$

where the allowed time $\text{window}$ is set to a 1 day offset. Finally, for all determined matched $(I - E)$ pairs denoted as $M$, we calculate the average daily ranking score as:

$$\text{AverageRank} = \frac{\sum_{(I-E) \in M} rank(I)}{|M|},$$

[5] www.globalincidentmap.com

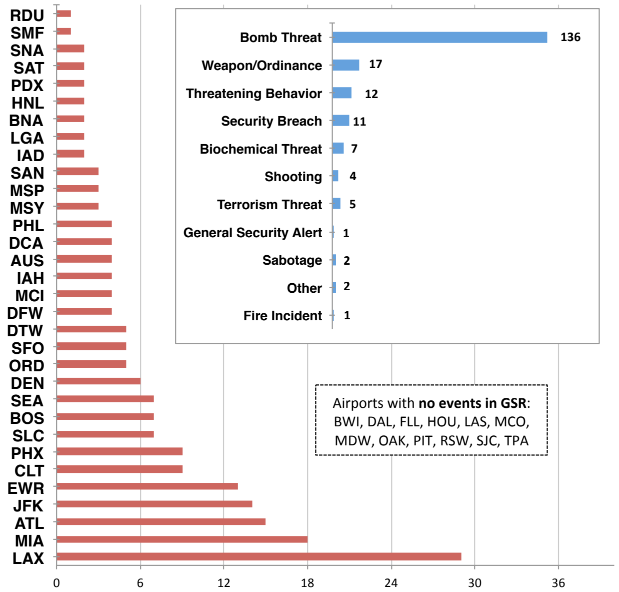


Figure 3: Distribution of 198 GSR events from January 2013 to August 2015.

where the $rank(\cdot)$ function returns the rank (by threat score) for the airport related with incident $I$.

**Rank Biased Overlap (M2):** We measure the similarity of determined rankings using the rank-biased overlap measure (RBO; (Webber, Moffat, and Zobel 2010)). The RBO measure is based on comparing the overlap of the two rankings at incrementally increasing depths. RBO is commonly used in comparing the results produced by online search engines and text retrieval systems.

**Ranking Loss (M3):** This measure follows from the observation that at any specific time, the ranking system may not provide the threat estimation for a specified airport. We use this metric to measure the coverage of our model.

**Ranking Baselines:** We make use of three baseline models to compare with our OSI approach - (1) Raw Twitter Volume (RTV) based threat rankings are generated from the total volume of tweets for each airport (larger volume translates to a higher rank for an airport); (2) DQE Volume (DQEV) refers to using only theme-relevant tweets and; (3) Tweet Score Value (TSV) using aggregated threat score (as shown in Equation 5).

Table 2: Comparison of ranking performance and relevance.

| Model | Average Ranking | Ranking Loss | Weekly | Monthly | Quarterly |
|---|---|---|---|---|---|
| RTV | 8.09 | 5 | 0.46 | 0.54 | 0.68 |
| DQEV | 4.73 | 66 | 0.41 | 0.55 | 0.70 |
| TSV | 4.73 | 66 | 0.37 | 0.49 | 0.68 |
| OSI (P1) | **3.50** | 66 | 0.38 | 0.40 | 0.69 |
| OSI (P2) | 7.80 | **1** | 0.49 | 0.54 | 0.69 |

## Ranking Performance & Relevance

We use $M1$ and $M3$ to evaluate overall ranking performance. In comparison to baselines, as shown in Table 2, PI does provide improvements in average rankingsa but P2 produces better (lower) ranking loss. We also evaluate the DQE approach for theme targeting compared to a fixed keyword set filtering (RTV). For each set of matched $I - E$ pairs generated from DQEV and RTV, we manually inspected tweets, to validate if the detected incident(s) correspond to the matched GSR event ($E$) or not. Table. 3 summarizes the results of this analysis, which shows that DQE is not only effective in reducing noise (false positive matches) but also achieves comparable recall (0.72) of ground truth events. To measure the relevance, we compare the similarity of rankings between our method and the GSR using the RBO score. As shown in Table 2, our quarterly performance across all airports is around 69% overall.

Table 3: Comparing of matching performance.

| Model | Detected Incidents | Matched I-E Pairs | True Positive Matches | False Positive Matches |
|---|---|---|---|---|
| DQEV | 8209 | 142 | 91 | 51 |
| RTV | 19757 | 184 | 89 | 95 |

## Case Studies

We visualize the ranked airports using rankflows which are an intuitive way to interpret progression (of *threat flows*) and encoding of relative rankings. We present our results in Figure 4, aggregated for every two months. Due to space constraints, we show only those airports that have appeared at least once in the top 15 positions of our rankings. We observed some very interesting patterns; for example, we can see characteristic (upward) bumps in rank flows for airports signaling threat incidents. For example in the case of Hartsfield-Jackson Atlanta International Airport (ATL), the rapid increase in threat rankings towards the end of August 2014 can be attributed to the fact that during this period two very disruptive events took place. First, Ebola screenings at airports were started in October 2014 and ATL was one of the 5 participating U.S. airports. Second, a gun smuggling racket was uncovered between ATL and JFK airports in late December 2014 (see Figure 5(b)), where an ATL baggage handler was arrested on December 10 in possession of 18 firearms [6].

As shown in Figure 5(a), our system reported a series of upticks in threat assessment for Detroit Metropolitan Airport (DTW), Los Angeles International Airport (LAX), and Denver International Airport (DEN) that coincides with evacuation incidents due to false bomb threats. In the last few years we have also seen cases where users on social media issue threats to airlines or specific flights. One of the most bizarre cases of such social media abuse was when a series of online threats were made using Twitter to over 20 different U.S. passenger planes in the last two weeks of January 2015. We

[6] tinyurl.com/hlksja2

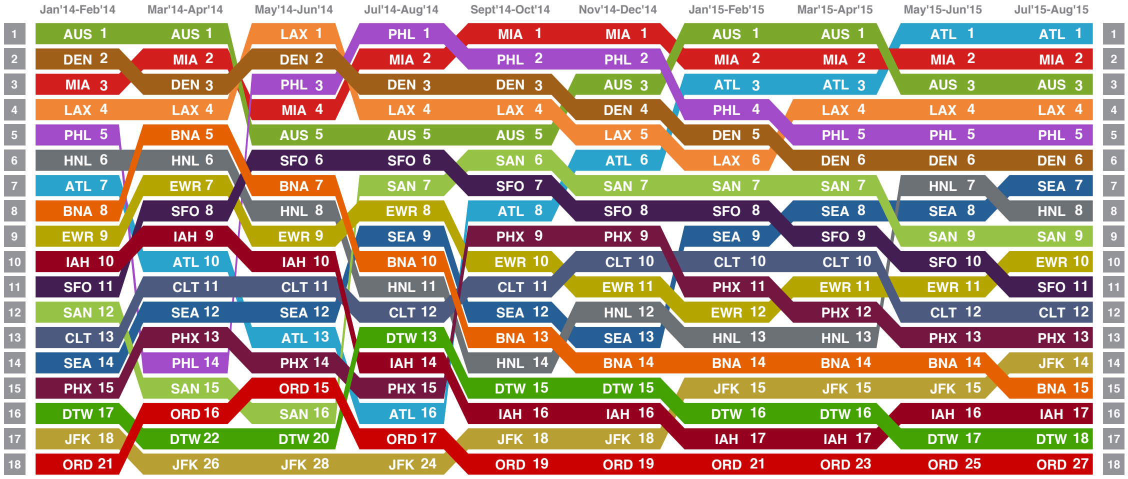


Figure 4: This rankflow digram indicates the relative rankings generated using our OSI (P2) model. The row order of the (threat) flows indicates how these airports compared among each other in their threat ranking (in grey box) and the number (rank) next to each airport code indicates their global ranking across all 45 airports (selected for the study).

noticed high threat scores (shown in Figure 5(c)) and corresponding *bumps* in relative ranks for several airports (such as SEA, JFK, and ATL), where flights were grounded and later evacuated in response to the online threats made by terrorist organizations (e.g., ISIS) through fake Twitter accounts.

## Related Work

With the growing popularity of Web 2.0 technologies, analysts from the Department of Homeland Security who have traditionaly relied on multi-agent based decision-support systems (Weiss 2008) for airport security, have realized the need for a class of automated systems that help provide situational awareness using open source indicators such as social media. The approach presented here is akin to other real-time, targeted theme tracking in social media which involves detecting the emergence and continuous tracking of the evolutionary dynamics of a specified theme such as crime (Wang, Gerber, and Brown 2012), disease outbreaks (Signorini, Segre, and Polgreen 2011), socio-political events such as civil unrest (Ramakrishnan et al. 2014), and in terrorism informatics (Cheong and Lee 2011).

The majority of existing research in targetted theme tracking adopts a classification framework to extract themes or latent topics from text as contextual features (Sakaki, Okazaki, and Matsuo 2010) or online social interactions (Lin et al. 2011). In contrast, our work leverages unsupervised methods (Zhao et al. 2014) that expand vocabularies and leverages on the reciprocal relationship between social media and conventional news. By combining these two mediums we are able to evaluate the trustworthiness of content in discovered themes on social media. News-Twitter reciprocity has also been shown to provide improved retrieval and relevance ranking of both tweets (Krestel et al. 2015) and news articles (Shuai, Liu, and Bollen 2012).

## Conclusion

Under the homeland security paradigm, tracking threats are increasingly important because any real world threat to general public or critical infrastructures such as airports, can be related to activities online. Since aviation security decision-making is a multi-dimensional process, we argue that the detected threats or reported incidents alone should not predominate the decision-making process. Threat-based rankings are valuable and should be used for priority-setting in conjunction with other factors, including expert opinion, public input, and the emergence of future risks.

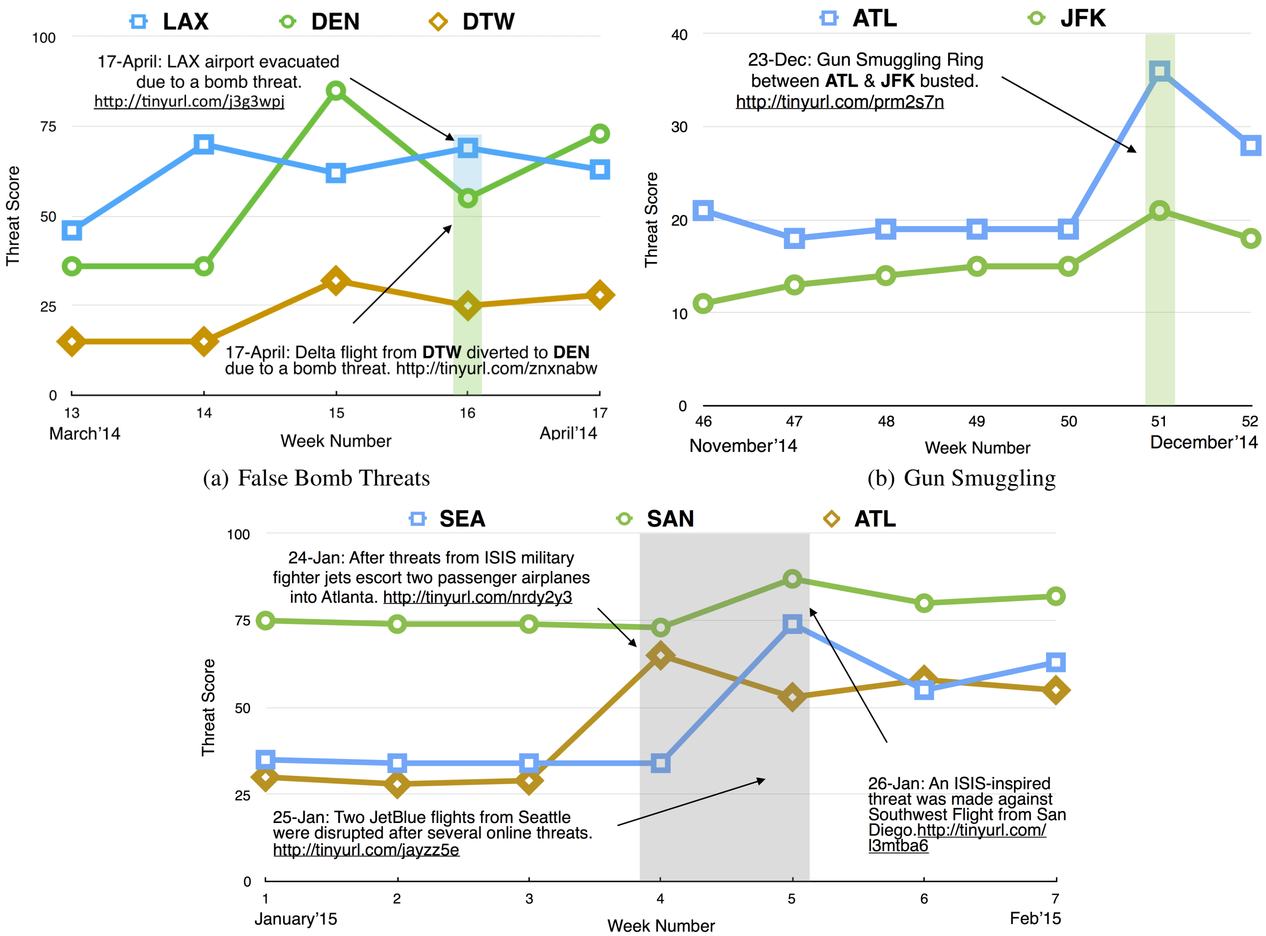
In this paper we have presented an automated system for tracking and relative ranking of airport threats that integrates real-time situational awareness using open source indicators (Twitter and news). Our evaluation over 45 U.S. airports illustrates the capabilities of our system in effectively identifying relative threats and emergent patterns. Our results should be interpreted as order of magnitude indications of potential threats, rather than actual predictions of attack incidence.

Future work is targeted at factoring semantic context and user's attitudes towards the security assessment of airports, so that comparative threat assessment and airport rankings are not solely model-driven but incorporate expert judgments and values.

## References

Cheong, M., and Lee, V. C. 2011. A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via twitter. *Information Systems Frontiers* 13(1):45–59.

Ester, M.; Kriegel, H.; Sander, J.; and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD '96*, 226–231.



17-April: LAX airport evacuated due to a bomb threat. http://tinyurl.com/j3g3wpj

17-April: Delta flight from **DTW** diverted to **DEN** due to a bomb threat. http://tinyurl.com/znxnabw

(a) False Bomb Threats

23-Dec: Gun Smuggling Ring between **ATL** & **JFK** busted. http://tinyurl.com/prm2s7n

(b) Gun Smuggling

24-Jan: After threats from ISIS military fighter jets escort two passenger airplanes into Atlanta. http://tinyurl.com/nrdy2y3

25-Jan: Two JetBlue flights from Seattle were disrupted after several online threats. http://tinyurl.com/jayzz5e

26-Jan: An ISIS-inspired threat was made against Southwest Flight from San Diego.http://tinyurl.com/l3mtba6

(c) Online Threats

Figure 5: These plots show timelines of several reported GSR events and our threat evaluation at those airports.

Krestel, R.; Werkmeister, T.; Wiradarma, T. P.; and Kasneci, G. 2015. Tweet-recommender: Finding relevant tweets for news articles. In *WWW '15*, 53–54.

Lin, C. X.; Mei, Q.; Han, J.; Jiang, Y.; and Danilevsky, M. 2011. The joint inference of topic diffusion and evolution in social communities. In *ICDM '11*, 378–387.

Llorens, H.; Derczynski, L.; Gaizauskas, R. J.; and Saquete, E. 2012. TIMEN: an open temporal expression normalisation resource. In *LREC '12*, 3044–3051.

Ning, Y.; Muthiah, S.; Tandon, R.; and Ramakrishnan, N. 2015. Uncovering news-twitter reciprocity via interaction patterns. In *ASONAM '15*, 1–8.

Ramakrishnan, N.; Butler, P.; Muthiah, S.; Self, N.; Khandpur, R.; Saraf, P.; Wang, W.; Cadena, J.; Vullikanti, A.; Korkmaz, G.; et al. 2014. 'beating the news' with embers: forecasting civil unrest using open source indicators. In *KDD '14*, 1799–1808.

Ramo, A. 1994. California comparative risk project (1994). toward the 21st century: Planning for the protection of california's environment. *Golden Gate University School of Law Digital Commons*.

Sakaki, T.; Okazaki, M.; and Matsuo, Y. 2010. Earthquake shakes twitter users: real-time event detection by social sensors. In *WWW '10*, 851–860.

Shuai, X.; Liu, X.; and Bollen, J. 2012. Improving news ranking by community tweets. In *WWW '12*, 1227–1232.

Signorini, A.; Segre, A. M.; and Polgreen, P. M. 2011. The use of twitter to track levels of disease activity and public concern in the us during the influenza a h1n1 pandemic. *PLoS one* 6(5):e19467.

Wang, X.; Gerber, M. S.; and Brown, D. E. 2012. Automatic crime prediction using events extracted from twitter posts. In *SBP '12*. Springer. 231–238.

Webber, W.; Moffat, A.; and Zobel, J. 2010. A similarity measure for indefinite rankings. *ACM Transactions on Information Systems (TOIS)* 28(4):20.

Weiss, W. E. 2008. Dynamic security: An agent-based model for airport defense. In *WSC '08*, 1320–1325.

Zhao, L.; Chen, F.; Dai, J.; Hua, T.; Lu, C.; and Ramakrishnan, N. 2014. Unsupervised spatial event detection in targeted domains with applications to civil unrest modeling. *PLoS ONE* 9(10):e110206.