CS224W Final Project Report: Uncovering the Global Terrorism Network

Julia Alison*
jalison@stanford.edu
Stanford University
Stanford, CA

Li Deng*
dengl11@stanford.edu
Stanford University
Stanford, CA

Zheqing (Bill) Zhu *
zheqzhu@stanford.edu
Stanford University
Stanford, CA

1. Introduction

One of the foremost security concerns facing the United States and the international community is the threat of global terror. Academics and governments increasingly seek to understand the underlying dynamics of the organizations behind the attacks.

As we describe below in our literature review, most of the analysis of terrorist groups has applied social network analysis to known organizations and performed empirical studies following major attacks. We have sought to approach the problem by investigating the dynamics of terrorist organizations in relation to one another as well as to develop an algorithm for predicting and preventing future attacks.

Section 2 includes a review of the relevant literature applying network analysis to understanding terrorist organizations as well as literature on some of the algorithms that we will employ in our study. Section 3 describes the Global Terrorism Database, our main dataset for the project, as well as other datasets that will supplement our work. Section 4 describes our methods for constructing networks to understand the international terrorist landscape, including the theoretical background for our methods and some pseudocode for our implementation. In Section 5, we develop methods for predicting future terrorist attacks with two approaches. Our SIS with Neural Network Model successfully predicted most of the major city terrorist attacks and among predicted cities, 54% were actually attacked by terrorism. Finally, we conclude our reports with reflections and suggestions for future academic work.

2. Literature Review

2.1. Theoretical Analysis

Existing literature has applied network analysis on the theoretical relationships between terrorist organizations and individual terrorists.

Everton (2009) [1], Clauset et al. (2008) [2], and Moon et al. (2007) [3] offer three distinct approaches to the theoretical study of terrorist network dynamics.

Everton (2009) criticizes the significant application of social network analysis to the study of terrorist networks by arguing that the focus on identifying and targeting key players within the network is misplaced. He seeks to understand the specific hierarchical structure within a terrorist organization (centralized leadership, decentralized, etc.) to help security forces target specific actors in a centralized network to disrupt its effectiveness. He identifies leadership as highly connected nodes within the network as well as an understanding of the cosmopolitanism of the network via measurements of the local clustering coefficients and average path lengths. His explanation is limited to individual organizations, but we are interested in some of his measures as applied to the global network.

Clauset et al. (2008) further supplemented our understanding of the underlying network structure of terror cells by building a hierarchal structure of a terrorist network via a maximum likelihood model to infer connections between nodes. They compared this model to a true terrorist network and found the two networks similar along metrics like average clustering coefficient and SCC size.

Moon et al.'s (2007) built a meta-structure of tasks and the agents assigned to complete them in order to infer the players who would have had contact in order to carry out the attack. The paper then applies social network analysis techniques by calculating betweenness centrality and total degree centrality on nodes (labeled as particular figures in the organization) for each of the models they built to identify key players within each theoretical organizational structure. The paper's primary strength is in its innovative approach to understanding why edges between actors within the terrorist network exist. However, like Everton, its techniques are limited to studying single organizations, while our paper takes a global approach.

2.2. Case Study

In addition to theoretical analyses that provide foundations to depict the underlying structure of terrorism net-

*Joint first author

works, researchers have conducted case studies that provide insights in individual terrorist organizations. Belli et al. [4] provided results from a case study that explores the social network of internal members of the "Hammound Enterprise", which was involved in trade diversion in order to finance terrorism organizations in Michigan. They found out that in these groups, members are highly interconnected, making organizations more efficient while vulnerable to detection. With key player analysis, three ringleaders are detected with a few secondary leaders, most of whom are Islamist extremists, which shows the property of an idea-centric organization. These analyses are very representative across many case studies we have found.

Krebs 2002 [5] is widely referenced in literature about terrorist network analysis. Following the September 11 attacks, the author used publicly available information about the 19 hijackers to construct a network of weak and strong ties based on the nature of their relationships with one another. By computing the degree distribution, betweenness, and closeness of the nodes, this paper depicts a sketch of the covert network behind the scenes, and allows him to identify the clear leader among the hijackers. Also from the analysis on clustering coefficient(0.4), and average path length among the nodes(4.75), the authors found out that this covert network "trade efficiency for secrecy" by remaining quite small and operating with little outside assistance.

Krebs' paper is one of the foundational documents in applying network analysis to terrorist organizations post-9/11. Because his research was conducted after the fact on a well-publicized case, he is able to delve into the nature of connections among actors in considerable depth, which lends credence to his findings about this specific group. However, his work does not lend itself to application on the larger terrorist landscape, because in most cases, information about the trust and familial connections among actors in a covert network is difficult to come by and verify. Thus, the insights gained from particular case studies can inform some of our approach to larger networks (i.e. identifying key players based on centrality, looking for hubs in a sparse network, etc.), but their approach is limited in scope and application to future study.

Instead of exploring the internal structures of each group, a global picture of the interaction between different groups may provide crucial linkage that connects pieces of small cluster networks. For instance, if we can establish a concrete theory of the collaboration between ISIS and extremists groups within the U.S., we can potentially figure out many sources of resources of terrorist attacks in the U.S. Another benefit of studying the global group network of terrorism is that it helps understand the global shift of terrorism distribution and predict future moves.

Though understanding individual organizations has importance, especially in the context of extremely destructive groups like al Qaeda or ISIS, we feel that the existing literature insufficiently accounts for the connections between terrorist organizations. Understanding those relationships can be crucial in inferring and predicting the future of terrorist organizations and behaviors, as we understand some organizations as behaving similarly or emulating central organizations to determine their next actions.

3. Global Terrorism Database

Global Terrorism Database(GTD)[6] is an open-source database including information on terrorist events around the world from 1970 through 2016. It contains information about 170350 terrorist events, and a total of 135 attributes for each event, including exact date, location, group, weapon, casualty and so on.

4. Methods, Algorithms and Evaluation

We first define a measure to value the severity of an event:

$$Severity_e = \alpha * N_{death} + N_{wounded} \tag{1}$$

where α is a parameter indicating how many wounds equivalent to one death in terms of severity. Here we use $\alpha = 3$ for our analysis.

We define the **lethality** of a terrorist group to be the sum of the severity of its all events:

$$Lethality_g = \sum_e Severity_e, \tag{2}$$

Then we compute the lethalities for all groups, and here are the top 10:

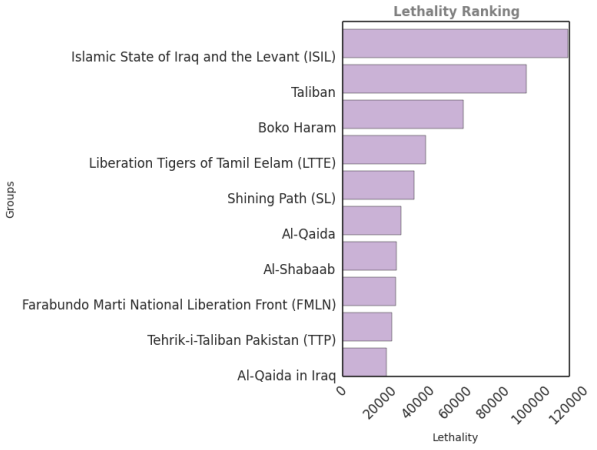

Figure 1: Lethality Rank

For simplicity, all figures below will show analysis result for the top 10 groups.

4.1. Relation Network Analysis

The GDT dataset contains a ”related” column for each event, indicating related events. Based off this information, we were able to construct the event relation network, where each node is an event, and an edge exists between two nodes if two events are related to each other.

We first conducted relation network analysis for USA:

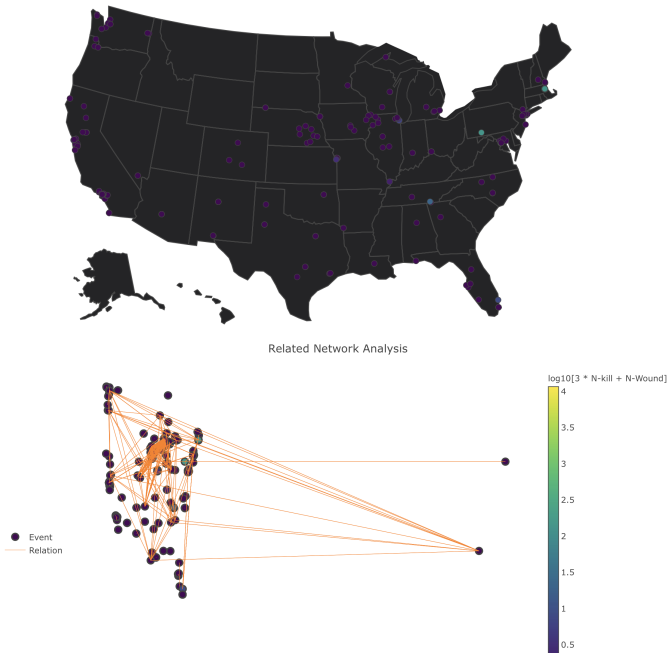

Figure 2: USA Related Terrorism

But we did not find interesting pattern for the relation network. Like Figure 3 shows, clustering coefficient across years seems pretty random, instead of an apparent pattern.

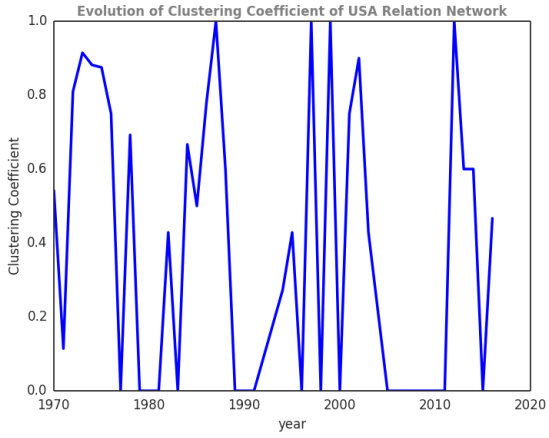

Figure 3: USA Clustering Coefficient Time Series

Then we switch our focus to relation network by groups, instead of countries. We plotted the heatmap for number of connections among events for each group from 1970 to 2016, to visualize the evolution pattern for the top 10 terrorist groups in terms of lethality.

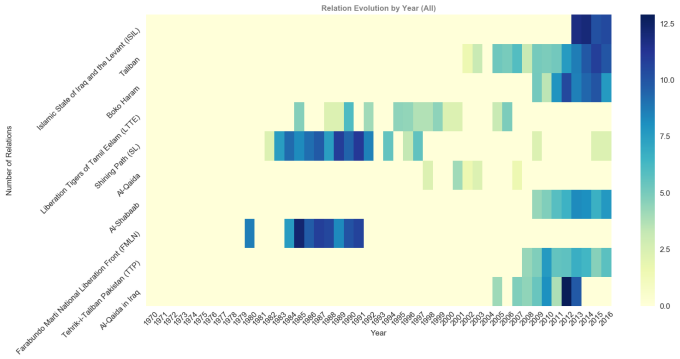

Figure 4: Evolution of Event Connections for Top 10 Groups

The related event may be caused by the same group, or a different group. And we define a relation connection be to ”external” if it is the latter case. Below is the evolution pattern for external relation connections:

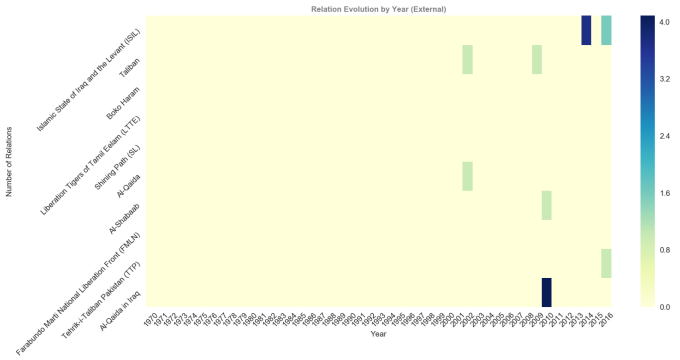

Figure 5: Evolution of External Event Connections for Top 10 Groups

As we can see, external connections constitute a tiny portion of all relation connections, i.e. most related terrorist events are actually caused by the same group, which presents one big challenge for us to uncover the terrorist network: there are few apparent connections between different terrorist groups. Interactions are hidden behind the scene, and we have to devise some other measure to expose group interactions. This leads to our next investigation into temporal evolution patterns of the groups.

4.2. Temporal Evolution Pattern

Our first approach is to analyze their temporal evolution pattern if terms of its lethality. From 6, we can build the evolution vector for each group: Let L_i be the lethality of

the group in year i, then each element in the evolution vector $E_i = L_i - L_{i-1}$ depicts the growth of the group in year i, i.e. the difference of lethalities of the group in two consecutive years: a positive value means the group is growing, while a negative value means the group is diminishing. So an lethality evolution vector captures the pattern of the evolution of the group.

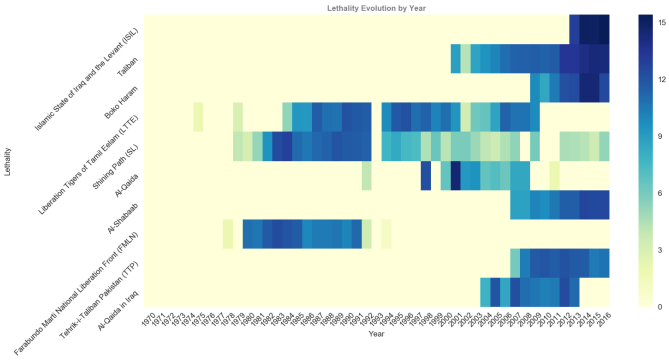

Figure 6: Evolution of External Event Connections for Top 10 Groups

Based on the intuition that if the two evolution vectors of groups are more strongly correlated, it is more likely that these two groups are connected somehow. So we construct the evolution vector for each group, and use *Seaborn clustermap* function to get the hierarchical clustering of the groups, based on the correlation of their evolution vectors. Pairs of groups clustered together indicate strong statistic correlation of their temporal evolution patterns.

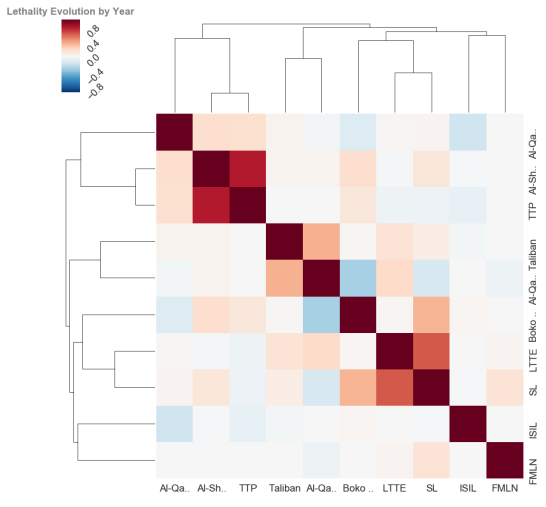

Figure 7: Clustering Based on Evolution Vectors

4.3. Similarity Network Model

Motivated by the temporal evolution vector approach where an evolution vector is constructed to capture the evolution pattern of a group, we came up with the next idea: construct a summary vector for each group based on its associated events in the GTD dataset, and compute the similarity matrix. Once we know the similarity score between any pair of groups, we may just set a minimal similarity threshold, and connect two groups with an edge if their similarity is above the threshold. So in this way, we can build a similarity network for all groups.

The summary vector for each group are based off all of its events in the GTD dataset. And we considering the following features for the summary vector:

- **Lethality**: lethality value defined in Equation 2;
- **Peak Year**: the year where the group caused the most events;
- **Attack Type**: the attack type that the group used the most in all of its events;
- **Weapon Type**: the weapon type that the group used the most in all of its events;
- **Target Type**: the target type that the group attacked the most in all of its events;
- $H2A$ **Ratio**: we define:

$$H2A = N_{\text{attacking home counter}}/N_{\text{all attacks}} \tag{3}$$

  which can be a measure indicating the group is more targeting its home country or foreign countries. The distribution of $H2A$ ratios for all groups is shown in Figure 8.

- **Longitude/Latitude**: mean value of the longitudes/latitudes of all its events; this can be a measure about the geographic active area of the group;

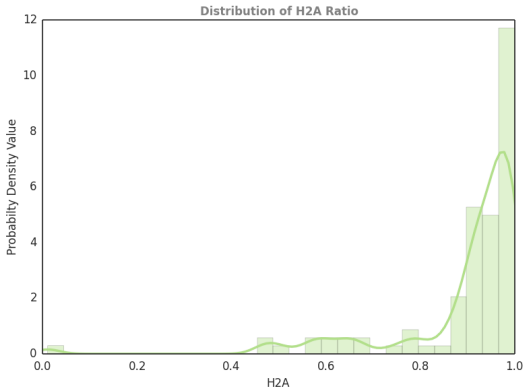

Figure 8: Distribution of $H2A$ Ratio for All Groups

We iterate over the full GTD dataset to get the group vectors, and come up with the following customized dot product of two group vectors to get their similarity value:

Algorithm 1 Similarity Network Construction

procedure SIMILARITYSCORE(v_1, v_2)
  $score \leftarrow 0$
  for $i = 1 : VectorDimension$ **do**
    if i^{th} feature is unbounded numeric value **then**
      $curr = min(v_1[i], v_2[i])/max(v_1[i], v_2[i])$
    if i^{th} feature is bounded numeric value **then**
      $curr = exp(-abs(v_1[i] - v_2[i])$
    if i^{th} feature is categorical value **then**
      $curr = 1[v_1[i] == v_2[i]]$
    if i^{th} feature is longitude/latitude **then**
      $distance =$ geographic distance by lon/lat
      $curr = exp^{-distance}$
    $score+ = curr$
  return $score$

procedure CONSTRUCTEDGES($g_1, g_2, threshold$)
  $v_1 \leftarrow GetGroupVector(g_1)$
  $v_2 \leftarrow GetGroupVector(g_2)$
  $similarity \leftarrow SimilarityScore(v_1, v_2)$
  if $similarity < threshold$ **then**
    return
  else
    construct an edge between g_1 **and** g_2

Here is the similarity matrix we have obtained for the groups:

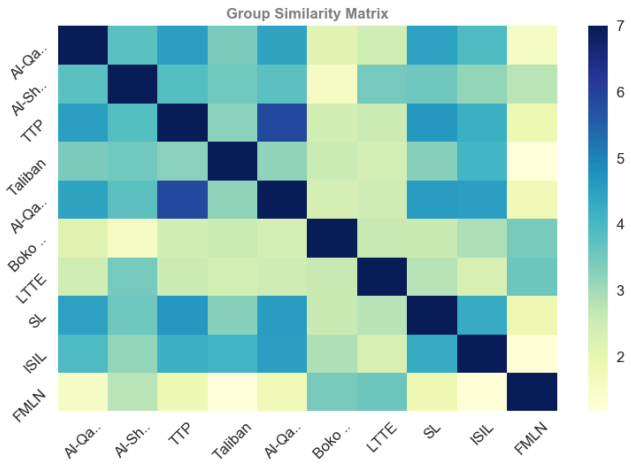

Figure 9: Group Similarity Matrix

As in Figure 9, darker colors indicate a stronger similarity.

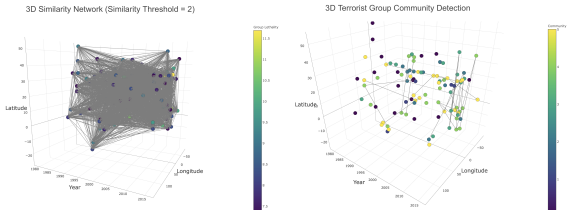

Figure 10: Similarity Network with $Threshold = 2$ and 5

The next step is to construct similarity network for groups based off their similarity score, and the edge construction process depends on our choice of the $threshold$ value.

As seen from Figure 10, too small threshold results in a too dense network, and too large threshold results in a too sparse network. After some explorations, we set the minimal threshold at 4.2, and we map all groups to a 3D graph, where x axis is longitude, y axis is year, z axis is latitude, see Figure 11.

Now with this similarity network, we use Girvan-Newman community detection algorithm provided by *snap* to identify communities. Figure 12 shows the 5 communities detected in the similarity network.

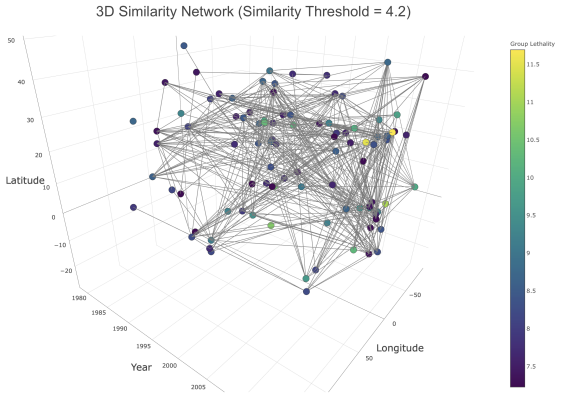

Figure 11: Community Detection in Similarity Network

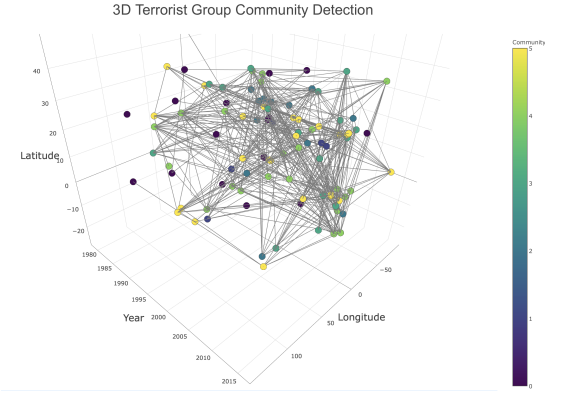

Figure 12: Group Similarity Matrix

Table 1: Communities

Community	Groups
1	**FARC, Ansar Allah, ISI**, FARC
2	**Boko ..**, Hutu .., Fulan.., MNR, CPI-Maoist
3	**Al-Sh..**, AQAP, Al-Nu.., Chech.., Donet..
4	**FMLN**, FDN, PKK, NPA, ARDE
5	**Taliban**, Khora.., South Africa, Narco..

Bolded groups are among the top 10 groups, to compare with previous clustering result.

There is no ground truth measure about the similarity values that we constructed, but we can validate our model by comparing the communities detected with the clustering in Fiture 7.

Here is the community clustering result from the similarity network:

4.4. Understanding Links Between Global Terrorist Organization Nodes

In trying to research possible algorithms for constructing edges between nodes, we found a base of literature on predicting underlying network dynamics by modeling cascade events and then predicting the edge configuration of k directed edges of influence most likely to result in the observed cascades.

In this section, we describe the networks between groups that we constructed using this algorithm and compare them to alternative algorithms used to calculate edges.

4.4.1 Hand-Labeled and Feature Similarity Networks

For each event, the GTD includes a hand-labeled column with the IDs of "Related" attacks as well as a list of all groups involved in each attack. To approximate a "ground truth" of terrorist group relationships, we constructed a graph with edges between all groups which appeared together in an attack or a "related" attack. As visualized below, with higher-degree nodes appearing with a bolder node coloring, the graph yielded several nodes with very few connections, and a few nodes involved in several attacks (usually in conjunction with other, similar groups).

The results of this more simplistic study yielded a result that was fairly in line with expectation, as demonstrated by the table below

Then, we constructed a graph based on attack profile similarity using a very similar metric to Algorithm 1 above (we will omit for the sake of avoiding redundancy).

Between the two methods, as described above, even with a very high similarity threshold, the attack-profile graph yielded an extremely dense result, while the related attacks were much more sparse. We compare them in the final subsection for this section.

Table 2: Top 10 Groups by Number of Relationships

Group Name	Connections
ISIL	34
Al Nusrah Front	29
Tehrik-i-Taliban	26
Ahrar al Sham	22
Lashker-e-Taiba	20
Lasher-e-Jhangvi	19
Shamiya Front	16
Taliban	15
Al-Assa Martyrs Brigade	15
Hizbul Mujahiden	14
Hamas	14

4.4.2 FastInf and NetInf

For this section, we treated terrorist events as a cascade of similar patterns (if one event were perpetrated successfully, we presumed that other groups might attempt to emulate the tactics, resulting in a pattern of similar attacks)

The database has information about the target, style of attack, and weapon used to carry out the attack, as well as its location. We used these features to label cascades of similar attacks temporally, which allowed us to implement two algorithms to infer the most probable edges between terrorist groups in the graph.

We begin with the simpler implementation, the FastInf algorithm by Alpay et al. [7], which considers each cascade a DAG constructed temporally (nodes point to nodes which adopt the pattern later). The algorithm then loops through each cascade to identify the maximum weighted spanning tree based on possible transmission.

For each potential edge, we consider the weight of that edge to be a measure of the probability that the destination node was infected in each cascade by the node immediately preceding it, assuming they were involved in this cascade. We assume that the probability of transmission decays exponentially (as other ideas might become en vogue, other groups exert influence, etc.)

Since edge distribution for the "related" network did not use a power law distribution, we elected to use an exponential transmission model.

$$w_c(i,j) = \frac{e^{time_j - time_i}}{\sum_{n_{i'}, t_{i'} < t_j} e^{time_j - time_i}}$$

Since the maximal weight will include all edges, we restricted the number of edges to explore to 500 to examine patterns among the most influential groups. Let E* be the optimal edges, and $W_{i,j}$ the sum of weights across all cascades for edge (i,j)

$$E^* = argmax_{|E| \leq k} \sum_{(i,j) \in E} W_{i,j}$$

The results are presented in the comparison section below.

Finally, we used the NetInf algorithm proposed in Gomez Rodriguez et. al 2012 [8] to infer edges from the cascade.

Net Inf similarly calculates an inferred edge list below a certain threshold based on the observed cascades. It considers all possible paths of propogation for the cascade (in this case attack technique).

If the probability of a cascade given a particular graph is the sum of the probability of a cascade given a tree times the probability of a Tree given a graph configuration:

$$P(c|G) = \sum_{T \in T_c(G)} P(c|T)P(T|G)$$

And the probability of the set of cascades is simply the product of the probability of each individual cascade (assuming conditional independence). Thus, we calculate our network as:

$$\hat{G} = argmax_{|G| \leq k} P(C|G)$$

This yields a similar simplified diffusion network influence problem:

$$\hat{G} = argmax_G F_C(G) \sum_{c \in C} max_{T \in T_C(G)} \sum_{(i,j) \in E_T} w_c(i,j)$$

NetInf is not a perfect solution, as it utilizes a greedy algorithm by picking the edge which adds maximum influence at each iteration. According to the paper, the algorithm achieves breakeven performance 90-99% of iterations on synthetic data with a known ground truth.

4.5. Comparisons

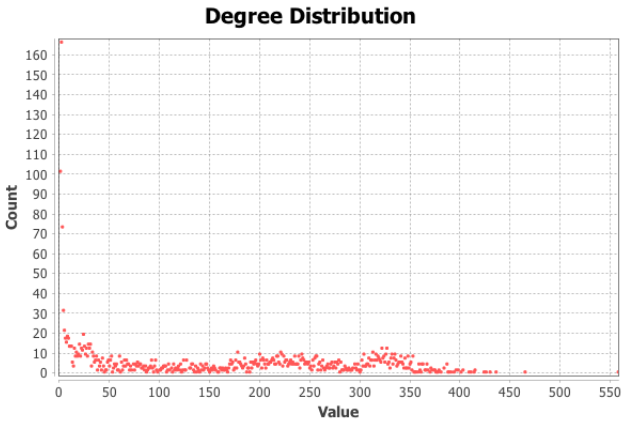

Figure 13: Degree Distributions for Computed Graphs-Similarity

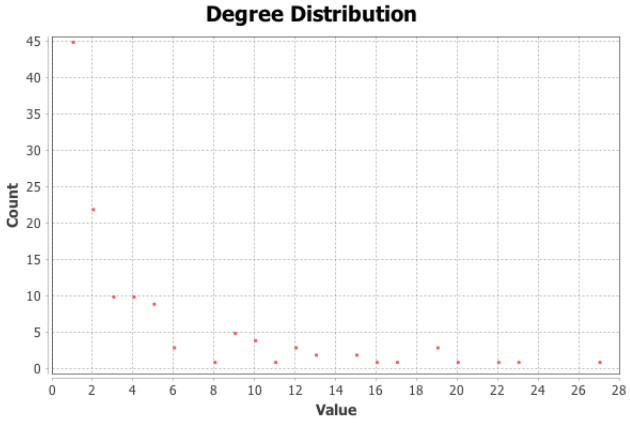

Figure 14: Degree Distributions for Computed Graphs-FastInf

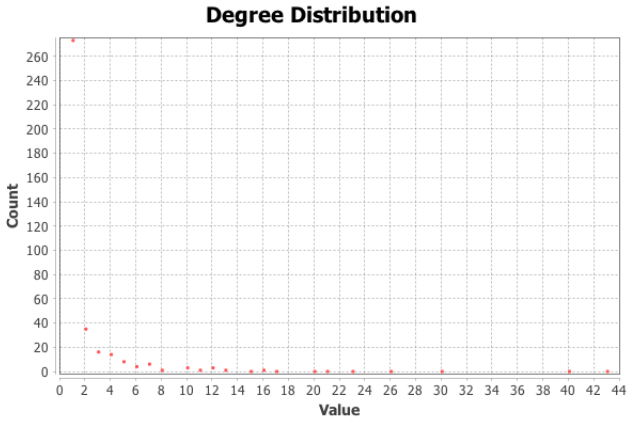

Figure 15: Degree Distributions for Computed Graphs-NetInf

Ultimately, none of the models tested yielded particularly heartening results. Because there is no ground truth global terrorism network, we could not compute precision or recall for our model. However, we could make observations which provide insight into the relevance of our generated networks.

As mentioned above, the "Related" network provided ten nodes of top degree which are almost constantly in the news, which is a heartening result, but ultimately does not add much in the way of novel information.

The similarity network was extremely dense, as many terrorist attacks utilize similar methods, especially in similar regions with similar targets during the same time period (i.e. similar tactics by suicide bombers in a particular country, similar tactics by leftists).

Finally, the two transmission models yielded much smaller graphs (most groups were disconnected nodes), which suggests that over time, few groups actively emulate others in their methods. Comparing this to the result above, it suggests that the time component that these models isolated was not particularly relevant, as groups attacking during a similar time period do not necessarily learn from a

Table 3: Comparison of Relationship, Similarity, and Inferred Networks

Graph	Nodes	Edges	Max WCC	Clstr Coeff
Related	3527	1644	1328	.0477
Similarity	2292	373942	373941	0.212
FastInf	245	300	292	0.0642
NetInf	388	500	356	0.040

single influential node, so their attacks can take place in any order, and the edges are not likely to be directed. The edges deemed most "influential" by these models were:

FastInf: Turks of Western Thrace, Tanzim, Shining Path, Popular Front for the Renaissance of the Central African Republic, and the Revolutionary Armed Forces of Columbia (FARC).

NetInf: Liberation, Tigers of Tamil Eelam, Black Liberation Force, July 14th Movement, Extreme Left, National Union for the Total Independence of Angola

From the data, these groups do not exhibit any particular relationship with each other. Because of the simplifying step in calculating the edge weights (only considered by time, not likelihood of transmission), the groups are disproportionately (though not entirely) groups that are older and fairly inactive. These results could be the result of a small sample size, in that our cascades were all fairly small (given the granularity of considering "inspiration" between attacks). Thus, these groups could all have sparked particular movements, and the results deserve further study.

Because the terrorism database is so expansive and diverse, we observe low clustering coefficients for all of our computed networks, as there are many classifications of activities which could be considered "terrorism" in the dataset, and the perpetrators are unlikely to be related in any measure. However, the large WCC in the similarity graph suggests that certain features of terrorism may be similar across groups.

5. Terrorism Event Prediction

Terrorism event prediction and location forecasting is an essential topic when analyzing terrorist event networks. Since most events occur on different dates, in order to make reasonable predictions with enough data, we discretize the event space across time into slices of events in each year. Given that data, we use the following two approaches to make forecasts of terrorist events regarding groups and cities. The first approach takes each individual group as an agent and the goal is to use techniques in motion tracking to see if the terrorist agents move with patterns. The second approach takes a global perspective and looks at the global terrorism network as an epidemic network with a Susceptible-Infected-Susceptible(SIS) model.

5.1. Particle Filter Based Event Tracking

One intuitive understanding of terrorism is that they spread. The typical way for terrorism to spread is that first the terrorist group would recruit radical people from their target locations and brainwash them. Then they start to use these people to either make terrorist attacks or recruit more people. From the procedure, we can observe that the spread is directional, which suggests that the spread of event could be tracked by a motion tracking system. From the theoretical point of view, we can consider the case where each terrorism group has its own movement transition matrix. In our approach, we adapt particle filter to model the movement of several terrorist groups to test our hypothesis of directionality.

In this problem, we treat the globe as a 180×90 grid where 180 represents latitude and 90 represents longitude. We have latitude and longitude data from each terrorist event as well as the terrorist group behind it. This gives us the opportunity to track the location and movement of each specific group with motion tracking techniques.

In order to make any predictions, we need to first define our metric. Since our terrorism events are defined in terms of earth coordinates, we introduce Vincenty distance s (ellipsoidal distance) [9]. Note that a is the length of semi-major axis of the ellipsoid, b is the length of semi-minor axis of the ellipsoid and σ is the arc length between points on the auxiliary sphere.

$$
\begin{aligned}
s &= bA(\sigma - \Delta\sigma) \\
A &= 1 + \frac{u^2}{16384}(4096 + u^2(-768 + u^2(320 - 175u^2))) \\
u^2 &= cos^2\alpha(\frac{a^2 - b^2}{b^2}) \\
\Delta\sigma &= Bsin\sigma(cos(2\sigma_m) + \frac{1}{4}B(cos\sigma(-1 + 2cos^2(2\sigma_m)) \\
&\quad - \frac{B}{6}cos(2\sigma_m)(-3 + 4sin^2\sigma)(-3 + 4cos^2(2\sigma_m)))) \\
B &= \frac{u^2}{1024}(256 + u^2(-128 + u^2(74 - 47u^2))),
\end{aligned}
\tag{4}
$$

Using a particle filter initialized uniformly across the earth grid defined above, we can track each individual terrorist group and predict their next move.

The particle filter generally gives a scattered sample of weighted particle and prediction like above for Taliban at their 100th event in the GTD database, see Figure 16. The black scattered dots are particles and the red dot is the predicted location. This is shown as an instance particle filter prediction. The x axis is latitude and the y axis is longitude.

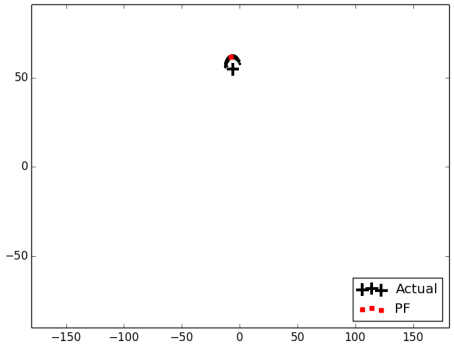

Figure 16: Taliban Particle Filter Sample

We use particle filter to track a terrorist group by event location at each occurrence. What particle filter does is as the following: first, the particle filter initializes uniform distribution of particles all around the world (the latitude, longitude grid). Each time when an event from the same group occurs, the algorithm updates the weight of each particle. Finally, when the algorithm is called to predict the next event, the particle filter calculates the weighted expected particle from all particles scattered. Particle filter is specifically used here due to the suspected non-linearity and non-Gaussian attributes of terrorism movements.[10]

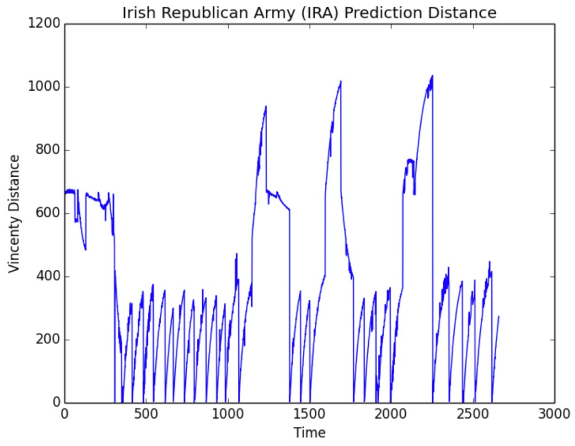

Figure 17: IRA Error Vincenty Distance (miles)

Figure 17 is an example of the error time series of particle filter prediction in terms of Vincenty Distance. Using particle filters provides us with an average of 532.37 miles in Vincenty Distance for predictions of location of Irish Republican Army events and 688.93 miles in Vincenty distance for ISIS events.

5.2. SIS Model for Terrorism

Motion tracking is a very micro-scopic approach to tackle terrorism. Given the spreading nature of terrorism, we can simply treat the occurrence of terrorism as outbreaks of epidemics.

One intuitive way of connecting the event network is through traffic network. As a preliminary step, we connect all cities that ever have terrorism events with airline route data from Openflight. Using time sliced network, we treat terrorism as an epidemic and follows an SIS model. We use data from year 1970 to 2015 as training data to produce a model to predict cities $S \to I$ and cities $I \to S$.

Given that we do not have the dynamics of city interactions, we introduce two neural networks for function approximation to classify the transition dynamics of $S \to I$ and $I \to S$ respectively.

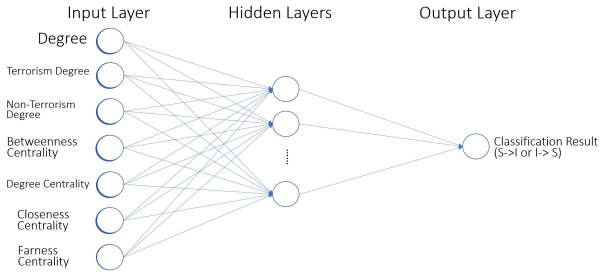

Figure 18: Neural Network Architecture

Figure 18 presents the architecture of the neural networks. At each time step, we feed in a city's degree, number of susceptible neighbor cities, number of infected neighbor cities, betweenness centrality, degree centrality, closeness centrality and farness centrality to a multilayer perceptron network for each instance of transition $S \to I$ and $I \to S$ to the respective neural network as training data.

Using the 2015 susceptible set of cities and infected set of cities, we make the following prediction on 2016 terrorism locations, where the size of each dot is proportional to the confidence level of prediction, see Figure 19.

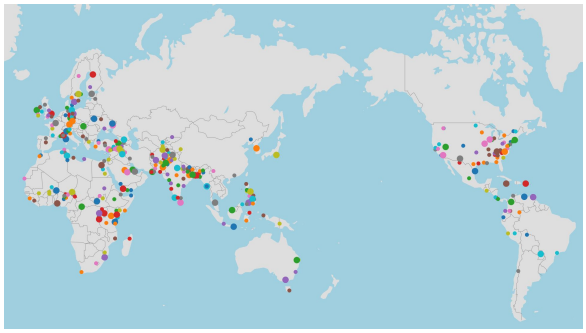

Figure 19: Global Terrorism Location Prediction 2016

The following is the ground truth of terrorism locations in 2016, see Figure 20.

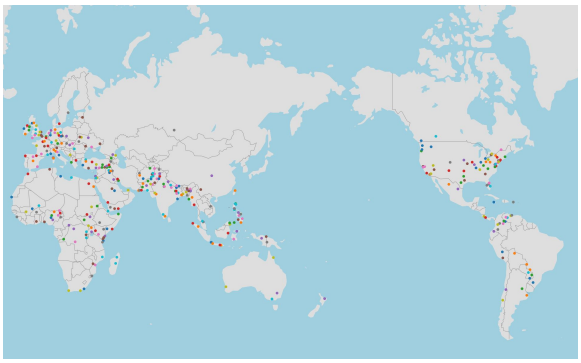

Figure 20: Global Terrorism Location 2016 (Ground Truth)

From the figures above, we can observe a clear matching pattern between predicted locations and ground truth locations. Among all the predictions, impressively, the SIS with Neural Network model was able to predict several major terrorism events in **London, Berlin, Montreal, Hamburg, Dublin, Copenhagen, Stockholm, Madrid, Paris, Milan, Istanbul, Florencia, Moscow, Bankok, Richmond, Washington D.C., San Jose and Toronto**. More importantly, among the **332** predicted cities, **54%** are actually attacked by terrorism events. Note that the total number of cities attacked by terrorism in 2016 according to GTD is **356**.

6. Discussion

In this project, we first try to uncover terrorist communities, and interactions among groups, by analyzing the relation network, and temporal evolution patterns. We proposed a simple algorithm for constructing the similarity network, based off the group vectors from GTD event dataset, and a customized similarity scoring function. Community detection algorithm is then applied on this similarity network for identifying underlying group interactions.

The first two sections involved coming to understand the network dynamics of a previously unexplored network of events, which attempted to understand the dynamics underlying the relationships between the groups behind the attacks. We found that a network could be computed among groups to find that there were many unifying qualities among them.

However, the results from the inference models suggest that we still have not found an accurate inference model for the edges between groups over the period.

With insights getting from our first two steps, our prediction has shown very promising results with traffic network and the SIS Model. Although predictions have shown significant impact, we have several future projects extend our current approach.

One thing that could potentially be the downside of our current approach is that neither Particle Filter nor SIS with Neural Network model provides substantial time constraints on our predictions, which leads to less actionable items for authorities to secure these areas. To resolve this problem, the potential solution is to narrow down the range of terrorist groups in a certain areas and their Modus Operandi. This would provide us with more features of how to classify with sparser feature sets.

Another aspect of our current approach is that the underlying network we are relying on is only the traffic network. One could easily One could easily imagine that current terrorism relies greatly on web and social network. Hence, tackling social network for terrorism is also a promising direction of predicting terrorism activities.

References

[1] F. Everton, Sean. Network topography, key players and terrorist networks, 2009.

[2] Aaron Clauset, Cristopher Moore, and Mark EJ Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 2008.

[3] Alexander H. Levis Il-Chul Moon, Kathleen M. Carley. Vulnerability assessment on adversarial organization: Unifying command and control structure analysis and social network analysis. *EEE Intelligent Systems, Special issue on Special issue on Social Computing*, 22(5):40–49, 2007.

[4] Roberta Belli, Joshua D. Freilich, Steven M. Chermak, and Katherine Boyd. Exploring the u.s. crime-terror nexus: Terrorist networks and trade diversion. *National Consortium for the Study of Terrorism and Responses to Terrorism*, 2014.

[5] Valdis Krebs. Uncloaking terrorist networks. *First Monday*, 7(4), 2002.

[6] Gary LaFree and Laura Dugan. Introducing the global terrorism database. *Terrorism and Political Violence*, 19(2):181–204, 2007.

[7] Altan Alpay, Deniz Demir, and Jie Yang. Fastinf: A fast algorithm to infer social networks from cascades. *CS 224W Final Project*, 2011.

[8] Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. *ACM Trans. Knowl. Discov. Data*, 5(4), 2012.

[9] T Vincenty. The meridional distance problem for desk computers. *Survey Review*, 21(161):136–140, 1971.

[10] Pierre Del Moral. Non-linear filtering: interacting particle resolution. *Markov processes and related fields*, 2(4):555–581, 1996.