Investigating Terrorist Attack Networks

Matthew Marcel, Punit Khiyara, Katherine Gibbons

To begin, we have often used network analysis to analyze our relationships and social ties with those close to us and within our extended social circles. Through the analysis of these relationships, we often find numerous clusters and critical nodes (certain people) within the clusters that act as the epicenter of our relationship with the rest of the nodes within the cluster. Hence, we began to wonder if it would be possible to uncover similar relationships amongst terrorist attack networks where each node represents a terrorist that participated in a certain attack or bombing within a country. If multiple terrorist networks spanning numerous countries and bombings over the years were to be entered into a network modeling and analysis program, we hypothesize that we would be able to uncover certain connections or relationships between multiple attacks that would have otherwise not been discovered through the individual analysis of each attack's network. Therefore, we set out to investigate terrorist attacks in the hopes of finding certain relationships between one another in order to determine areas of vulnerability within the network that could be exploited by government agencies.

In order to analyze and search for relationships between multiple terrorist events, we had to collect data on multiple terrorist networks that were preferably related to major attacks. As such, we found multiple crime network datasets through the UCINET software website that is provided to students by the university's school of behavioral sciences' criminal justice department. Using this website, we were able to find a couple of interesting datasets for which we traced back their original source to a website called John Jay & ARTIS Transnational Terrorism Database (Link to Datasets' Source). This database consisted of 12 terrorist network datasets complete with 24 separate CSV files for nodes and relationships between nodes for each attack. The nodes in the dataset each represent an individual terrorist that took part in the corresponding terrorist attack.

The challenges we faced with this project was understanding the extensive codebook behind all of these attacks. Luckily, all of the attacks we investigated used the same codebook (Link to NODE Codebook) (Link to RELATIONS Codebook). We chose to only use nine out of the twelve attacks from the database since three attack network datasets did not follow the standardized column structure implemented by the rest of the attack network datasets. The columns of interest to us from the relations datasets consisted of the ID, Tie_ID, Tie Year, Tie Extinguish, Kinship, 1985_1989, and 1990_1994. From the node datasets, the columns we were interested in consisted of ID, Group, Color, Arrest Date, Release Date, Death Date, 1985_1989, and 1990_1994. While we primarily focused on the ID and Tie_ID columns for the purposes of importing the datasets into Gephi, we chose to keep the date and time series columns in the dataset as well for potential future use. Additionally, we required the columns across each terrorist attack network’s dataset to be identical so that we could concatenate several datasets using a Python Jupyter notebook. We used Bali Bombings 2002, Bali Bombings 2005, Australian Embassy Bombing, S.E.A. Attack, Christmas Eve Bombings, Hamburg 9/11, Madrid Train Bombing, Philippines Ambassador Residence Bombing, and the Vivace Bombing. In terms of scope, our terrorist attack network datasets range between the years 2000 and 2005.

Another challenge we faced consisted of importing the datasets into Gephi. Since each terrorist attack had two datasets (nodes and relations), we uploaded 18 datasets into Gephi. Prior to doing so, however, we realized that we needed to change the columns “ID” to “Source” and “Tie_ID” to “Target” from the relations dataset for each attack in order to help Gephi understand the datasets and the relationships between their nodes. Additionally, when uploading each attack, we needed to upload the node file to the node section and the relation file to the edges section for each attack. Then we would run the modularity statistic on the network for color and network diameter statistic to output the betweenness centrality of each node. Once we ran these network statistics, we colored each node and edge based on their respective modularity class and sized the nodes based on their betweenness centrality. Finally, we also had to upload the master dataset, consisting of all 9 attacks, by following the steps previously stated.

The last challenge we had with understanding the data completely was the fact that some of the individuals were double-coded throughout the nine different bombing attacks (especially with the Indonesian attacks). This meant that one node (individual) was coded into multiple different datasets, and this became a challenge when adding additional attack labels to the nodes for the node master dataset. To be clear, the fact that some of the nodes were duplicates throughout the attack datasets clued us to the fact that some of the attacks were related. The issue that arose was how we were going to label the nodes in the master network for Gephi (as each label is based on the dataset of origin). Gephi, fortunately, automatically deleted the duplicate nodes within our dataset and applied the first attack label the node was related to.

Social Network theory is a sociological theory that attempts to explain human behavior by studying their interactions and relationships with others. Criminal Justice has adopted this general theory and applied it to why people would commit crimes. In simple terms, social network theory in the criminology setting would explain that someone who has friends, families or acquaintances that engage in criminal behavior is more likely to engage in criminal behavior themselves. When we analyzed our dataset at the preliminary level, we saw that few of the nodes had kinship ties to other nodes. Therefore, we looked towards the social network theory to understand the strength of weak ties within a network -- an individual can and will most likely be influenced to engage in criminal behavior themselves if the person is surrounded by many weaker relationships with criminals (friends, friends of friends, and acquaintances).

Therefore, as we visualize our terrorists bombing datasets, not only are we trying to find patterns and relationships between the individuals (critical nodes and pinch points) but we are running the modularity statistic on these networks to see how these terrorist networks are being created and which clusters we can identify from within these networks. Our first step in analyzing the datasets through our Gephi visualizations was to visualize each of the nine bombing datasets as presented in Figure 1.

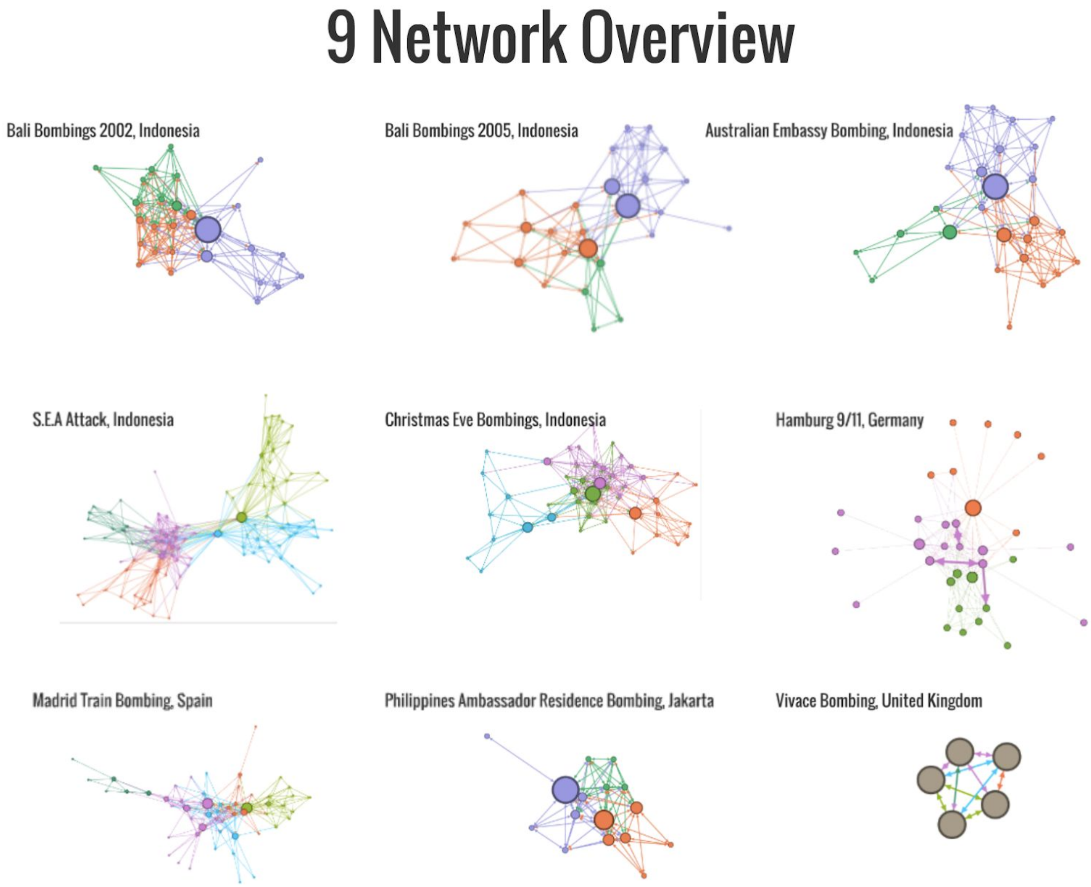

9 Network Overview
Bali Bombings 2002, Indonesia
Bali Bombings 2005, Indonesia
Australian Embassy Bombing, Indonesia
S.E.A Attack, Indonesia
Christmas Eve Bombings, Indonesia
Hamburg 9/11, Germany
Madrid Train Bombing, Spain
Philippines Ambassador Residence Bombing, Jakarta
Vivace Bombing, United Kingdom

Figure 1: An overview of the 9 terrorist networks that were analyzed

For our visualizations of the nine different bombings, we applied the modularity statistic, which helped us identify potential subnetworks and clusters within each bombing. Coloring the nodes based on the modularity classes allowed us to also see how closely related the nodes are within the network responsible for the bombing (notice that most of the visualizations have only three different modularity classes). Each node represents a person that is connected to the specific attack and the size of the node is based upon the node's betweenness centrality. Betweenness centrality defines a person's role in allowing information to pass throughout the network from one point to another which means that the bigger the node, the more information they are passing on to other individuals within the network. Therefore, we can identify critical nodes and pinch points in each network by finding the nodes that are much larger in size compared to the other nodes in that network.

When trying to take down a network, one wants to be able to identify the nodes that have the most relationships within that network because if you eliminate that specific node, a large part (if not the entire) network will collapse. As we look at the nine networks individually, we can see that certain networks would be easier to eliminate based on pinch points such as these four Indonesian attacks (Bali 2002, Austrialian Embassy, Bali 2005, Philippines Ambassador) and the London attack (Vivace). This is because those specific Indonesian attacks have one to three very large nodes that would cause nearly the whole terrorist network to collapse if eliminated. The London attack, unlike the rest of the bombing networks, actually does not vary in modularity class because there was no identifiable subnetwork within this bombing. To fully understand why, we colored the edges based on kinship instead of modularity for this attack alone, and found that all five individuals are family-related. Nonetheless, this network is still easier to collapse since there are only five nodes involved. The other networks have emerging complexity in comparison (S.E.A Attack, Christmas Eve Bombing, and the Madrid Train bombing) making it harder to identify the pinch points as there are too many links between all the nodes for one or two nodes to stick out in terms of size comparison. The next step in our network analysis was to combine all nine bombing datasets together and create one mass terrorist network that would include them all as presented in Figure 2 below. This would allow us to see if there are any relationships between the separate attacks.

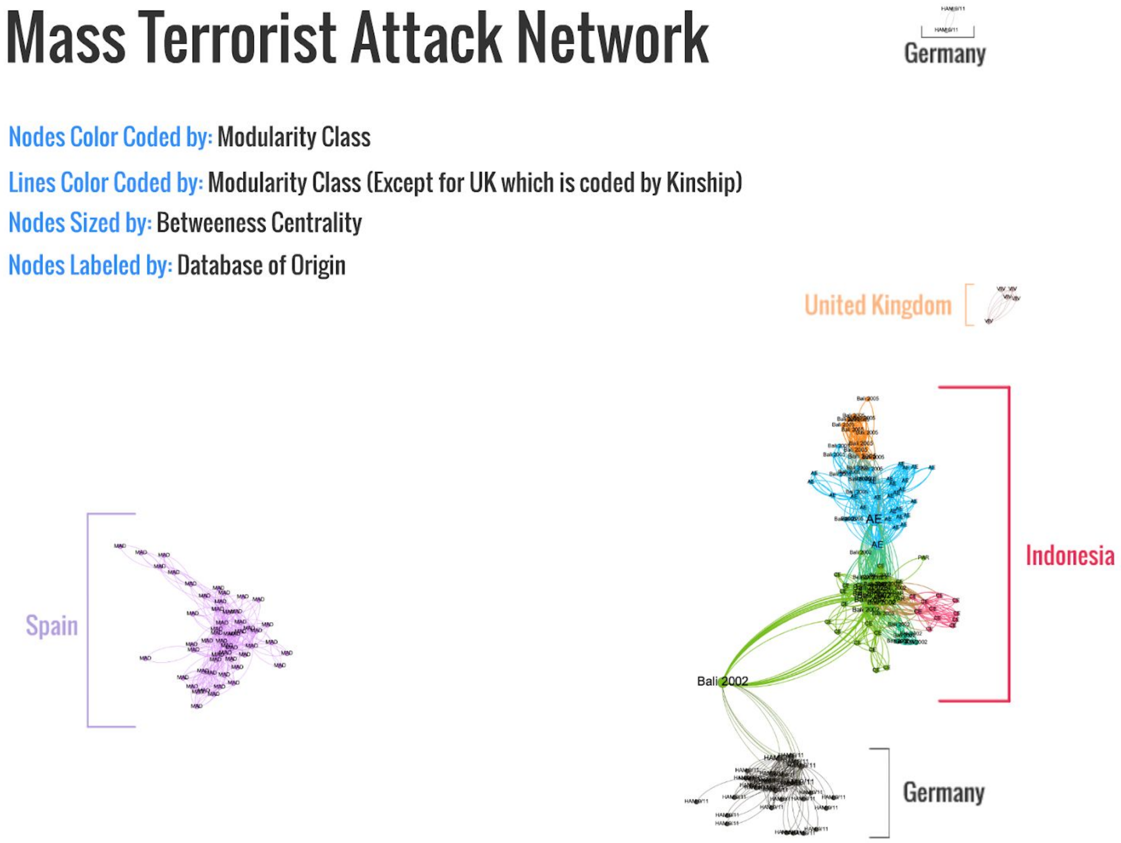
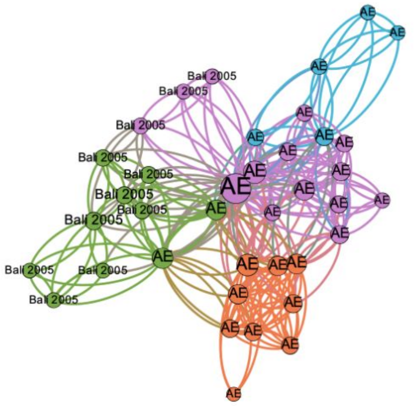

Mass Terrorist Attack Network
Nodes Color Coded by: Modularity Class
Lines Color Coded by: Modularity Class (Except for UK which is coded by Kinship)
Nodes Sized by: Betweeness Centrality
Nodes Labeled by: Database of Origin
Germany
United Kingdom
Indonesia
Spain
Bali 2002
AE
Germany

Figure 2: An overview of the Mass Terrorist Attack Network

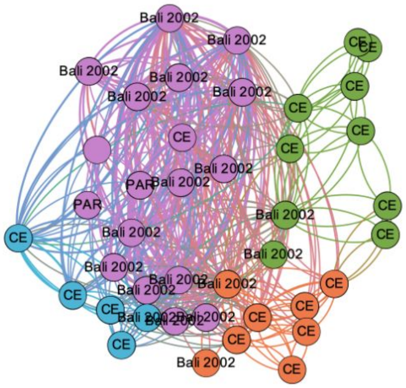
Above is the network that was created after running the Yifan Hu layout, running the modularity statistic for color-coding, and labeling each node with which attack they were originally related to. Right away, we can see that the bombing in Madrid (purple) and in the UK (tan) are not related to the other bombings and the networks responsible for those bombings are isolated from the rest of the data. Unlike the bombings that have taken place in Indonesia, we can see that each node is related to each other as we observe nodes that are identified with one bombing having a connection with other perpetrators of the same bombing. This helps us acknowledge that geography acts as a big factor in the connections between networks responsible for bombings since it is much easier to establish a terrorist network and carry out terrorist attacks if all the perpetrators are residing in one country.

Looking at the modularity classes in the Indonesia bombings (green, red, orange and blue), we can also see that there are subnetworks that primarily focused on one attack. For example, the blue nodes are primarily related to the Australian Embassy bombing that took place in 2004 but, they are still heavily connected to the 2002 and 2005 Bali bombings. This suggests that time is also a big factor to consider when analyzing the different clusters created within our network as our modularity classes seem to differentiate based on the time of the attack. However, the largest, most interesting (yet simultaneously terrifying) revelation we found within the combined network is that there was one individual that played a role in both the Hamburg Bombing (black nodes) and the Bali 2002 bombing (green nodes).

After this revelation, curiosity seemed to consume our group as we wanted to investigate how the clusters in our mass terrorist network were being formed, what they were composed of, and if there were any similar patterns within them that were noticeable in the larger network as well. In other words, we wanted to look at the subnetworks within the subnetwork.



Modularity Class 2



Modularity Class 0

Modularity class 2 is represented by the blue nodes and modularity class 0 is represented by the green nodes from the large network. Each network here is just one modularity class upon which we ran another modularity statistic to further investigate how the smaller network was being formed. From the modularity class 2 network, we can specifically observe how the individuals who had a role in the Bali 2005 bombing were linked to the Australian Embassy bombing perpetrators. Likewise, the modularity class 0 network shows us similar findings but between the Bali 2002 bombing and the Christmas Eve bombing.

Exactly what do all these networks mean? By developing several different networks and layers within those networks, we can see how certain individuals are related to each other, something the dataset alone would not show us (especially when most nodes have no kinship with other nodes). This is key to understand if you are the police or a government agency that needs to capture terrorists, weaken communication among terrorist groups, and breakdown large terrorist networks. We have created visuals that can help one target the individuals that have the largest influence within these networks, as well as find key relationships between bombings across Indonesia and Germany. The reason that we felt the finding that an individual had ties to bombings in two different countries was the most interesting revelation is because we can speculate that the individual was able to learn terrorist tactics from the Germany bombing (that took place in 2001) and apply it in Bali a year later. Moreover, through this revelation, we could speculate that there are some larger terrorist networks in the world that have the ability to share information with each other through such individuals. Lastly, our creation of the "sub-subnetworks" is a practice that can help a great deal when trying to take down a large, complex terrorist network. To be able to analyze a cluster deeper and deeper until you find pinch points and critical nodes will help police units eliminate multiple sections of the large terrorist network rapidly in comparison to looking at the cluster on the surface and not being able to identify the critical nodes. It's a way to apply a "chain of command" to the network in order to find out how the network expands upon itself.

Furthermore, we chose to create a Prezi to tell our terrorist network story. We wanted the ability to flow through the networks in a methodical way that mimics oneself interacting with the networks, especially since the PNGs from Gephi are not interactive. We began by explaining the data set, how we manipulated the dataset, and how we combined it into two separate master data sets (nodes and relations). We then wanted to show the separate networks of each attack we looked at. As such, we made 9 separate Gephi networks to analyze each dataset's network individually before we combined all 9 of the attacks together. We wanted to see the nodes colored by modularity class and sized by betweenness centrality for each network; it allowed us to view critical nodes and pinch points as well as the emerging complexity of certain networks. We also wanted to identify that one network, the Vivace UK bombing, had no modularity and instead each node within the network was connected based on kinship, indicating familial ties. Next, we move into the combined, master network that includes all nine different attacks. This shows the interconnectivity between the Indonesian attacks, Indonesian attack perpetrators linked to German attacks, and separate networks like Spain and the UK. We then zoom in to show how we ran modularity again (modularity on modularity) to potentially identify sub-networks within each modularity class' network for modularity classes 0 and 2.

Our data-driven story Prezi can be accessed through this link:
https://prezi.com/o_ap8e97sf_k/?utm_campaign=share&utm_medium=copy

As for the roles performed by our group members for the project, Matthew and Katherine acted as the network designers for the project and created various different networks based on the terrorist activity datasets using Gephi. Matthew used his knowledge from his criminal justice major, classes, and reached out to previous professors for additional information pertaining to this project such as finding credible, crime data sources. Additionally, as the chief social network analyst, Matthew was the first to draw conclusions about what we could conclude from the network analyses and the respective impact of these conclusions for government entities and combating modern terrorist networks. In order to use the datasets collected from the John Jay & ARTIS Transnational Terrorism Database, our data manipulations specialists Punit and Katherine wrangled the 18 nodes and relationship CSVs and found the attacks that we were able to use for our network analysis project based on column similarity. Katherine created the initial master datasets by using Python and Jupyter notebook to concatenate the 18 selected datasets from the database into 1 master node dataset and 1 master relations dataset consisting of data from all 9 major terrorist attack networks. Moreover, Punit rearranged the master node dataset to include labels for each node based on their respective attack's dataset, codename, city, and country. Additionally, Punit acted as our Prezi expert since he had an extensive level of prior experience using Prezi and therefore, he outlined, added, and formatted the content of the Prezi while opening it up to the group members for additional contributions.

Works Cited

- "Betweenness Centrality." Betweenness Centrality - an Overview | ScienceDirect Topics, https://www.sciencedirect.com/topics/computer-science/betweenness-centrality.
- "Network Data." John Jay & ARTIS Transnational Terrorism Database, http://doitapps.jjay.cuny.edu/jjatt/data.php.