SDSC 2102

Social Network Analysis

Project Report

World Terrorism Network Analysis

Instructor: Dr. QING Ke

MALHOTRA Akul (56698828)

BUDIMAN Darren Nathanael (56672641)

City University of Hong Kong

Semester B 2022/2023

1. Background

Terrorism is a global problem which has been persisting for a long-time affecting individuals, governments and general populations across the globe. Terrorism is a calculated use of violence to generate fear among a population and support a specific political agenda. It has historically been used by political groups, revolutionaries, extremist groups and even religious groups in some cases (Jenkins, 1998). While the world has found solutions to many of its problems through leveraging modern-day technology, there still isn't a direct answer to terrorism. Over the past few years, even though the number of deaths by terrorism has decreased, the number of terror attacks has surprisingly remained the same or increased. Considering the data for 2021, the number of deaths caused by terrorism went down by 1.3%, but the number of terror attacks increased (Vision Of Humanity, 2022). This is highly worrisome.

With the recent conflicts in Ukraine, it is expected that they shall drive a rise in cyber and traditional terrorism. In the Ukraine conflict of 2014, the state recorded over 60 official terror attacks with more than 3000 civilian casualties. (*Ukraine Russia Crisis: Terrorism Briefing - Ukraine*, 2022) The present conflict between Russia and Ukraine is even more widespread and thereby concerning. On the brighter side, many organizations and governments have started to leverage technology using historical data to counter the terrorism battle. (*Countering the Use of New and Emerging Technologies for Terrorist Purposes | Security Council - Counter-Terrorism Committee (CTC)*, n.d.)

The focus of this paper is similar, focusing on conducting a general analysis of countries affected by terrorism through network representations and network analysis and also an analysis which is more specific to Afghanistan to understand which cities in Afghanistan are being attacked together by terror groups. Network analysis permits in-depth analysis of relations among participating entities and further exploration of the social structures that emerge from these networks*. (Network Analysis| Science Direct, n.d.)* The paper shall try to discover insights about countries which are highly affected by terrorism and shall also take into account the terror groups which are responsible for such widespread attacks. Using this research organizations can have a deeper understanding of terror groups and also understand vulnerable countries/ cities which have been affected the most.

A link prediction analysis shall also be conducted specifically in Afghanistan to understand which cities in Afghanistan are vulnerable and have been attacked by the same organization. Terror groups have political agendas and target specific cities or provinces; thus, such analysis is crucial to know which cities are vulnerable and other cities which are linked to it. This shall be important as it can help Afghanistan to prepare in advance and prevent future attacks. It is expected that through this paper, governments and international organizations can have a better and well-informed understanding of this major problem and thereby take necessary steps to protect the people of their specific countries.

2. Methodology

A number of steps were followed to conduct the analysis ranging from data collection, data pre-processing, general analysis and link prediction based on the results collected in the initial steps. Different tools were utilized to conduct the project which includes Gephi for network visualization and analysis, Python Programming Language for Data Preparation and Preprocessing, Networkx library In Python for Analysis, Pandas Library for Data preparation, Microsoft Power BI for visualizations and Node2Vec library in Python for node embeddings.

2.1) Dataset

The Dataset, namely 'Global Terrorism Database' was taken from Kaggle, however, the data collected was very large with 181,691 rows and 135 columns. In addition, it was unstructured and was not in the right format to conduct social network analysis. Therefore, a number of steps were taken to pre-process the data and make it suitable for conducting further analysis. Firstly a number of different columns having low importance were removed. In this project, we used four main attributes to visualize our graphs:

1. ['iyear'] : This field contains the year in which the incident occurred.
2. ['country_txt'] : This field identifies the country or location where the incident occurred.
3. ['city'] : Name of the city, village, or town in which the incident occurred
4. ['gname'] : The name of the group that carried out the attack.

However, another problem arises as there are 82,782 terrorist attacks for which the attackers have not been identified. Hence, we decided to remove them from the data and considered them to be "Unidentified Attackers", this is because it is irrelevant to conduct analysis for terror groups which have not been recognized and shall take away the project's attention from well recognised terror groups, who can be identified and we can take specific action against whom.

2.2) Data Preprocessing & Visualization

Each visualization used a different data preparation technique.

Table 1
Main Terrorism Data

	country_txt	attackers
0	Dominican Republic	MANO-D, Dominican Popular Movement (MPD), Tony...
1	Mexico	23rd of September Communist League, Revolution...
2	United States	Black Nationalists, New Year's Gang, New Year'...
3	Uruguay	Tupamaros (Uruguay), Tupamaros (Uruguay), Tupa...
…	…	…
191	Vietnam	Provisional National Government of Vietnam, Pr...

Note. This table is the format of preprocessed data for our first visualization, it shows countries in the column "country_txt" and the terrorist group who attack the country in the column "attackers."

For the first visualization representing countries and the attacking groups, the purpose was to have a general overview of different attacking groups and the countries they attacked. The data was prepared accordingly and represented each country and all the organizations it was attacked by in the past. The data was then visualized using Gephi and the data was filtered according to Edge weight. This is to remove unnecessary edges or links since certain countries might have been attacked by an organization only a few number of times without any political agenda, however since we wish to conduct an analysis of countries being attacked by organizations for political motives, and take into account only necessary edges, so only the edges wherein the edge weight was higher than 2 were included.

Moving on to our second visualization for the project which puts a specific emphasis on Afghanistan. Afghanistan was chosen because of the high volume of edge weight because of the large number of attacks taking place in the country. Out of all countries Afghanistan had the highest weighted degree and out of all attackers, one of its attackers, Taliban, had the highest weighted degree.

Table 2

Terrorist Attacks Data in Afghanistan

	gname	victims
0	Black December	Kabul
1	Shia Muslim extremists	Kabul
2	Muslim Guerrillas	Ghazni, Herat
3	Hizb-I-Islami	Kabul, Kabul, Kabul, Kabul, Taloqan, Mazari Sh...
…	…	…
35	Jundallah (Pakistan)	Dashti Qala district

Note. This table shows our data after prep processing for the 2nd graph, it shows the terrorist group who attack the city in the column “gname” and the attacked city in the column “victims.” Furthermore, the format is remodeled for a victims-victims graph because it needs to connect from one victim to another. This format is used to make co-authorship graphs.

In this graph a link between 2 cities represented them being attacked by the same terror group. Data pre-processing for the graph was conducted using Python, wherein the resultant data represented the attacker in one column and the city/country in the other column. A pair of 2 cities basically symbolized them being attacked by the same organization. This visualization aimed to see which cities in Afghanistan are getting attacked by the same terror organization. Since terror organizations have political agendas and attack similar cities together, the graph shall be useful to visualize which cities are targeted more and together. Several subgraphs were created within this graph using partitioning based on different parameters. Degree distribution was also calculated using Networkx.

To conduct further analysis, the team also forecasted future linking of nodes using link prediction. This was conducted so that the analysis could be used in future and help organizations and governments prepare in advance for future such attacks.

2.3) Link Prediction

In order to allow the team's analysis to be further utilized by the Afghanistan government, the team went on to conduct link prediction to predict which cities are to be attacked by the same organization again in the future. Terror groups having specific political agendas attack specific cities and understanding which group of cities are more linked to the other could help us have a deeper understanding of the terror situation and also help organizations prepare well in advance. The task was done as a classification task, wherein the input data contained the two city names and a target variable which was 0 referring to no link and 1 referring to a link between two cities. In order to conduct the prediction, the Node2Vec algorithm was utilized which allows mapping nodes in a graph to an embedding space, which also preserves the network structure information. Thus this allows us to further conduct Machine Learning. This was done using the Node2Vec Library in Python, after which Logistic Regression was used to classify a link being present or not between the two cities, a link being present symbolized that they could be attacked by the same organization again in future. Due to computational power constraints the team could not implement SVM and Random Forest Classifiers, however the codes for the same have been attached as well.

A list of all Python-related code can be found here:

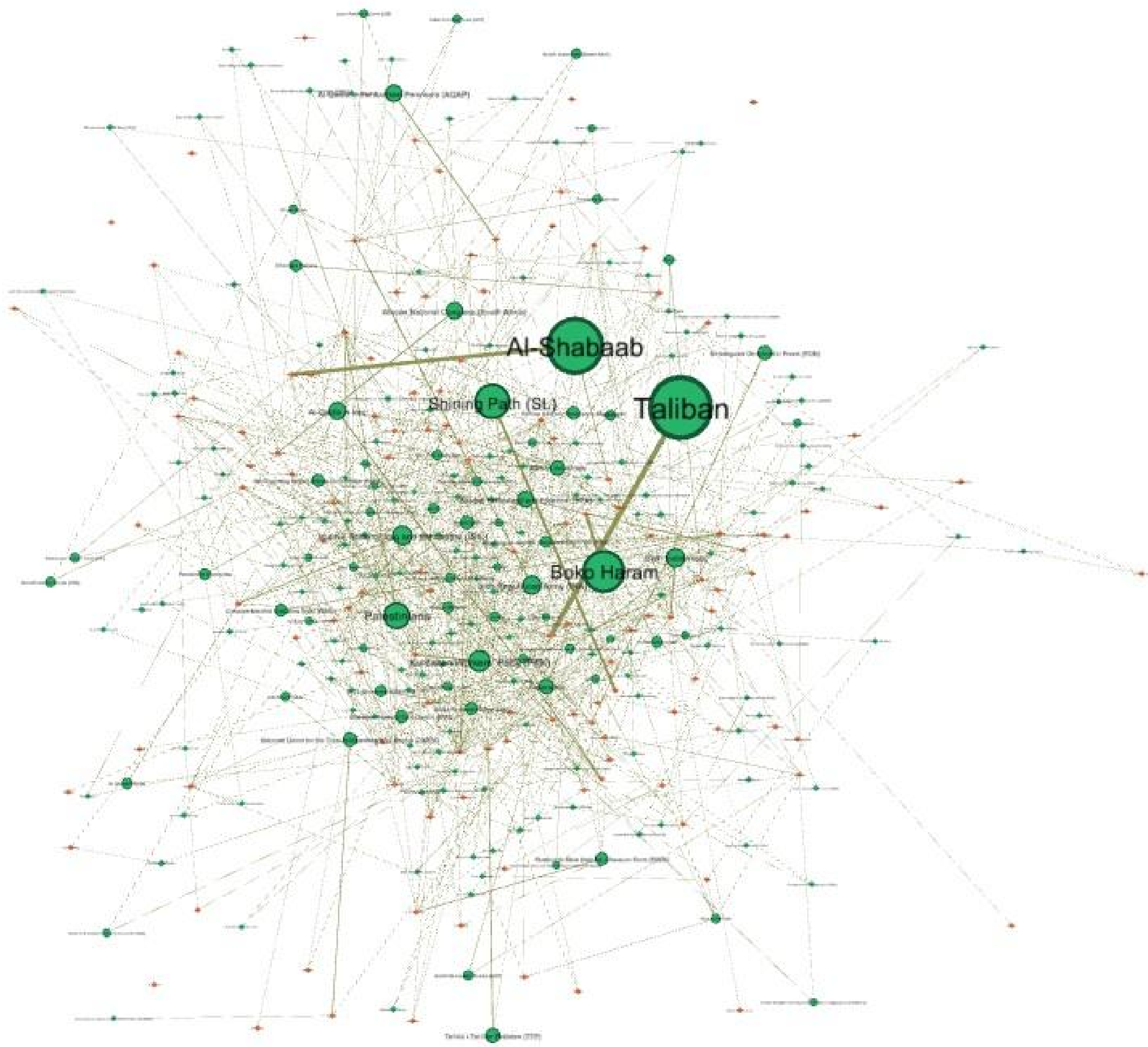
https://github.com/darrenbudiman/SDSC3016-SNA-DataPreprocessing

3. Data Analysis

3.1 Terrorism Graph

Figure 1

Terrorist Attacks Main Graph: Country-Attacks Graph



Note. The green nodes represent the terrorist groups while the orange nodes represent the country being attacked. The graph is undirected and weighted, the weight is based on the frequency of attack involving a country and an attacking group.

The Terrorist Attack Main graph has 351 nodes and 785 edges across the graph. Only 11.07% and 17.99% of nodes and edges are visible since the previous filtering process which results in a small number of both figures. Originally, there were 3171 nodes and 4364 edges. The data were filtered by the degree in the range of 2.0 - 99.0 and edge weight in the range of 2.0 - 3174.0 to cut down the irrelevant nodes, which represent countries that have been attacked by terror groups only a very few times.

3.1(a) Graph Density Analysis

The network was very dispersed, shown by a low graph density of only 0.01. The low density symbolizes how the number of edges are much lower than the maximum number of edges, meaning that all terror groups have not attacked most countries. This symbolizes how terror groups have specific political agendas and target specific countries suiting their specific political agenda and not all countries. The low density graph is actually good for a terrorist network.

3.1(b) Weighted Degree Analysis

We note that the average weighted degree is very high for the graph (123.7), meaning specific countries are attacked by specific terror groups a large number of times. The analysis of the weighted degree will be explained more in Table 3.

Table 3
Terrorist Attacks Main Data Sorted by Weighted Degree

Id	Frequency	Type	Weighted Degree	Degree	Number of Triangle	Eigenvector Centrality
Afghanistan	1	country	3355.0	25	0	0.3387
Taliban	3203	attackers	3203.0	3	0	0.0674
Al-Shabaab	2813	attackers	2813.0	6	0	0.0557
Somalia	1	country	2426.0	18	0	0.2548
Nigeria	1	country	2047.0	19	0	0.3009

High weighted degree means that for each connection in the graph, it occurs for around 124 times. More specifically, on average each country in the network was being attacked more than 120 times by each attacking group. However, the average weighted degree cannot be said to have significant information since the range between the maximum, and minimum weighted degree is large. Afghanistan has the highest weighted degree among all countries, followed by Somalia and Nigeria. Among all attackers, the Taliban has the highest weighted degree, followed by Al-Shabaab and Boko Haram. Based on the weighted degree analysis, Afghanistan was attacked by terrorist groups on 3355 occasions, of which 3174 of them happened to be from the Taliban Organization. The weighted degree analysis also helps us conclude that terrorists are attacking specific countries a large number of times, and countries like Afghanistan have a very high weighted degree because of the large volumes of attacks from terror groups.

3.1(c) Why does France have the Largest Degree but not the Largest Weighted Degree?

The following table displays the highest degrees for countries.

Table 4

Terrorist Attacks Main Data Sorted by Degree

Id	Frequency	Type	Weighted Degree	Degree	Number of Triangle	Eigenvector Centrality
France	1	country	657.0	58	0	0.8503
Lebanon	1	country	857.0	57	0	1.0
Italy	1	country	427.0	51	0	0.7965
Pakistan	1	country	727.0	50	0	0.5852
Israel	1	country	197.0	45	0	0.6850

This data had an average degree of 5.983, which means each country generally was being attacked by five to six different terrorist groups. France had the most with a total 58 degrees showing that it had been attacked by the most terrorist groups across the database. However, its weighted degree was only 657.0, relatively small compared to Afghanistan. If we divide the weighted degree by the degree, on average each terrorist group attacked France 12 times each.

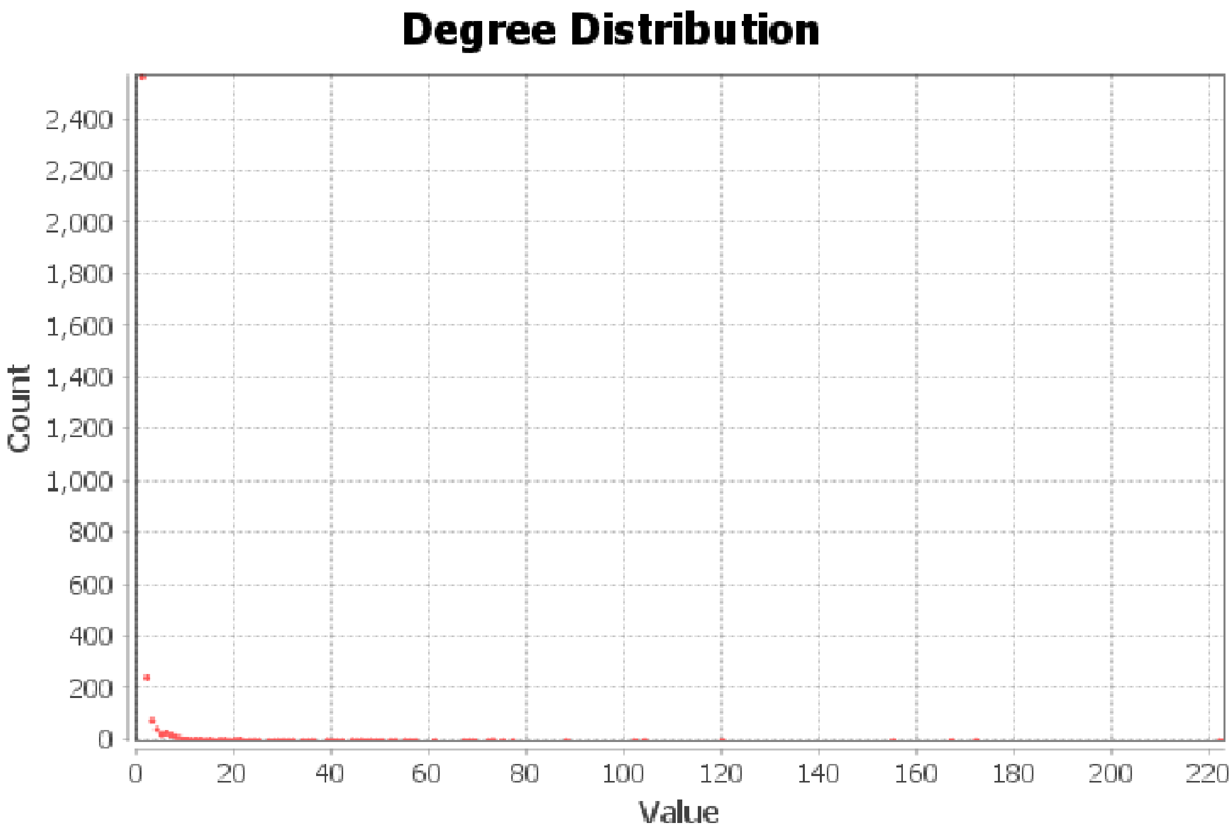
The degree graph shown in table 4 shows that France had lots of different organizations that tried to terror the country. However, the attack was not that frequent, as shown by the small number of weighted degrees they have. This is in contrast to Afghanistan wherein the number of attacks have been high but the degree is not so high. So France is a peculiar case wherein it is being attacked by many terror groups but the number of attacks are not many. This is highly concerning that so many terror groups are targeting France.

3.1(d) Degree Distribution Analysis

When a usual degree distribution is fitted, the graph turns out to be heavily tailed as shown below and thus is not useful to give sufficient information.

Figure 2

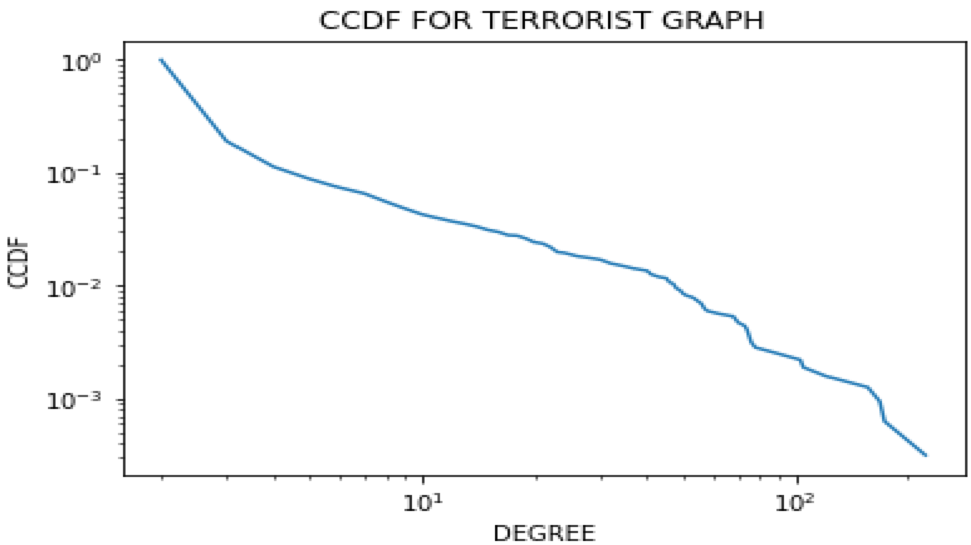
Terrorist Attacks Main Data Degree Distribution Created by Gephi



To create a better distribution, we fit the data with a CCDF in log-log scale.

Figure 3

Terrorist Attacks Main Data CCDF Degree Distribution

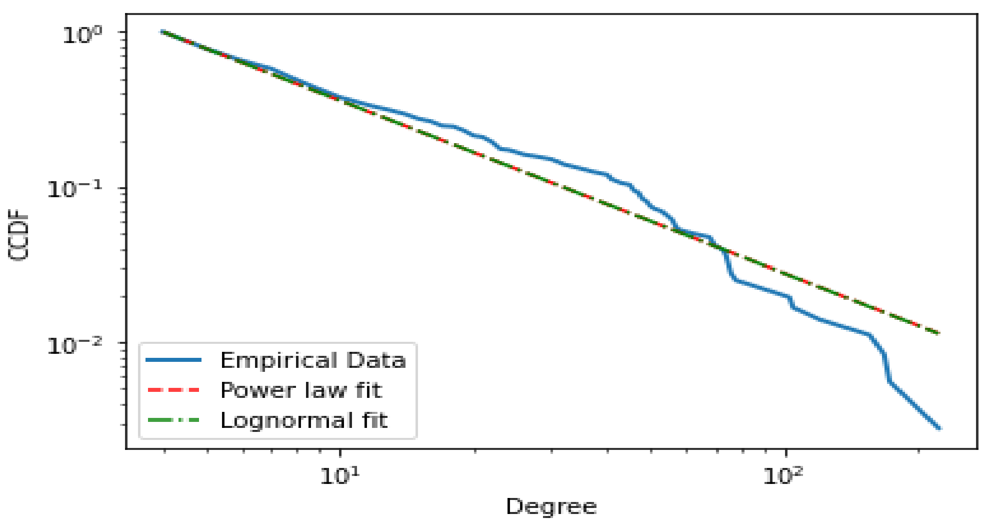


As we note, it is much easier to view the CCDF for the graph. It is interesting to see that close to 0.1 % of the nodes have a degree greater than 10 and a very small percentage has higher than 100. Thus the average degree of the graph is not very high.

It shall be useful to try to fit in some distributions. First, we shall try to fit in the Power Law and Log Normal Distribution first.

Figure 4

Terrorist Attacks Main Data Power Law and Log Normal Distribution



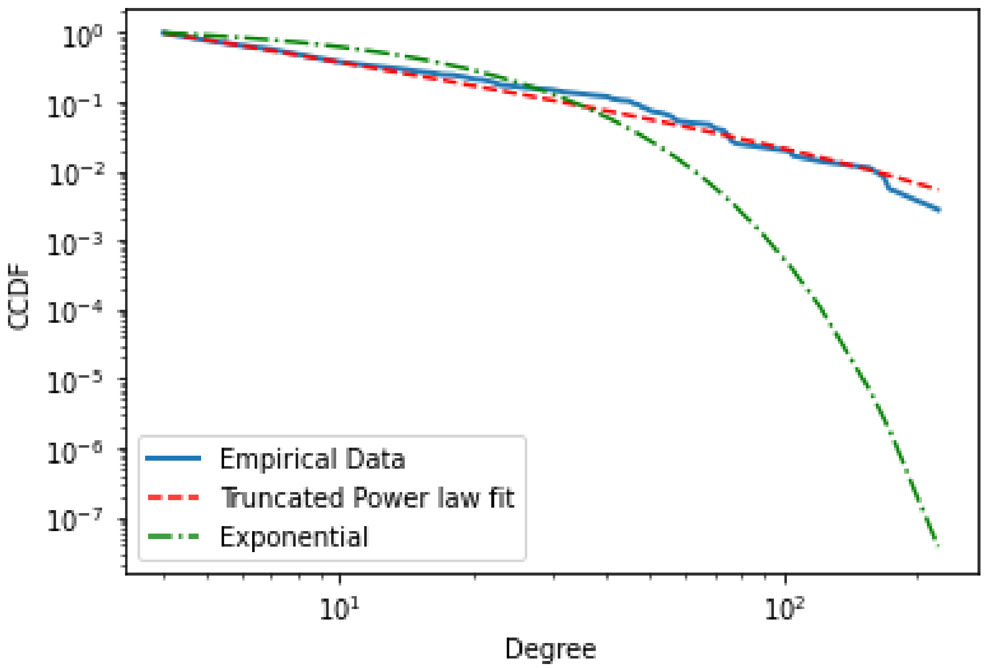
In figure 4, we can see that power law and lognormal were fitted into the degree distribution. It is interesting to see that the two distributions largely overlap with each other. On comparison between two distributions, R turned out to be 0.004 meaning power law is a better fit, however p value was significantly high, which is 0.85, thus the value of R is not significant and maybe due to statistical fluctuations.

A strict conclusion cannot be made from the above graph regarding which distribution is better.

So now, Truncated Power Law and Exponential distribution shall be tried.

Figure 5

Terrorist Attacks Main Data Truncated Power Law and Exponential Distribution



Truncated Power Law seems to fit the data well.

On comparison of Truncated Power Law's fit with other distributions here are the results,

Table 5

Terrorist Attacks Main Degree Distribution

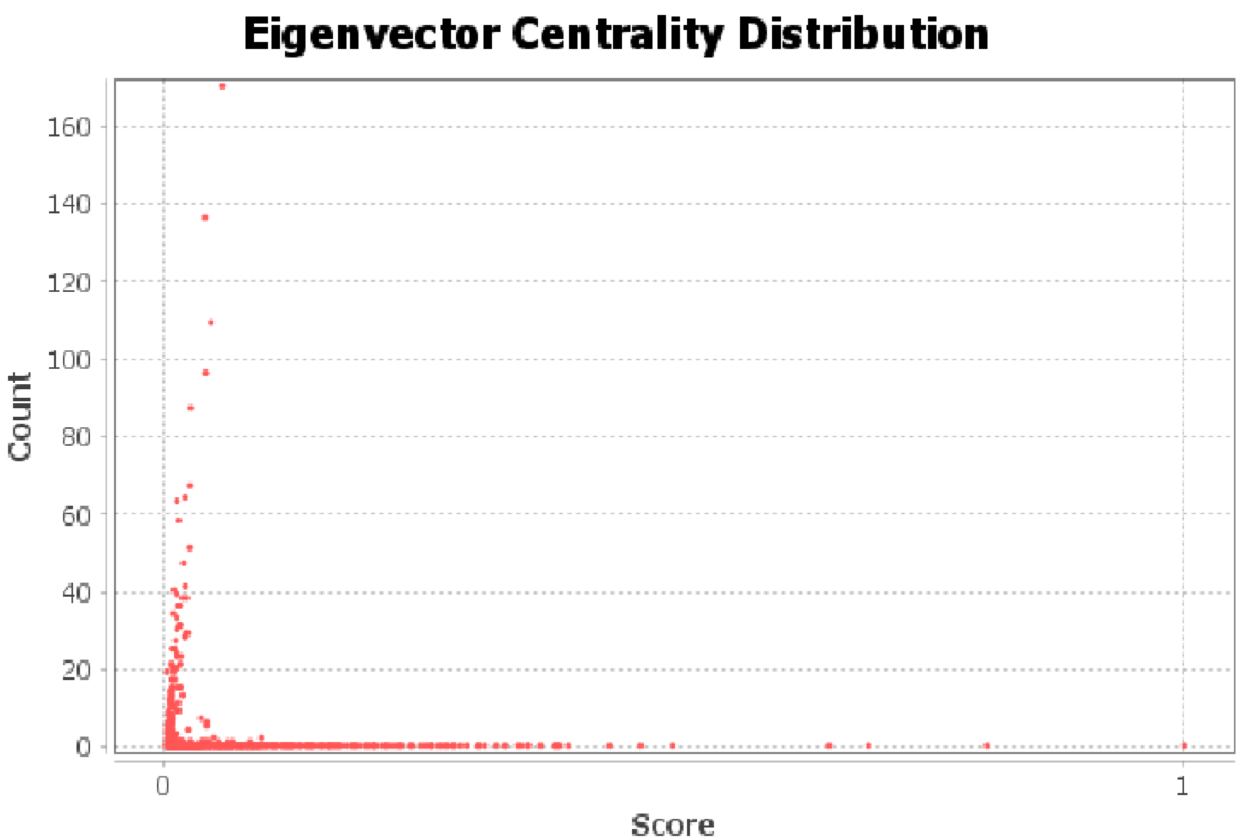
	R Value	P Value
Exponential	134.59	6.083924031868098e-12
Power Law	1.506	0.08
Log Normal	1.51	0.111

Thus, Truncated Power Law has a better fit than above distributions but since p-value is not less than 0.05 for Power Law and Log Normal, the significance is not very high.

3.1(e) Centrality Analysis + Why does Lebanon have the Highest Eigenvector Centrality?

Figure 6

Terrorist Attacks Main Data: Eigenvector Centrality Distribution



From the eigenvector centrality distribution graph, we can see that most of the nodes have close 0 eigenvector centrality. This fact means that most of the terror groups are small-scale and do not add significant effects to the data network, and it is also noticeable that even though most countries have centrality close to 0, there are a few countries having higher centrality scores close to one.

Table 6

Terrorist Attacks Main Data Sorted by Eigenvector Centrality

Id	Frequency	Type	Weighted Degree	Degree	Number of Triangle	Eigenvector Centrality
Lebanon	1	country	857.0	57	0	1.0
Muslim Extremists	487	attackers	486.0	44	0	0.9110
France	1	country	657.0	58	0	0.8503
Italy	1	country	427.0	51	0	0.7965
Black September	120	attackers	115.0	28	0	0.7172

It is interesting to see that Lebanon has the highest eigenvector centrality among all countries, also it is noticeable that France has a very high eigenvector centrality. Eigenvector centrality more simply measures how important a node's neighbors are in the network. France and Lebanon have huge degrees and are thus linked and being attacked by a large number of terror groups, which are important and add value to the entire network, thus Lebanon and France have higher eigenvector centralities, but since the total volume of attacks on these countries is not so high, their weighted degree is not extremely high. Even though the weighted degree is not so high for Lebanon and France, the number of groups attacking it are very high which is extremely concerning.

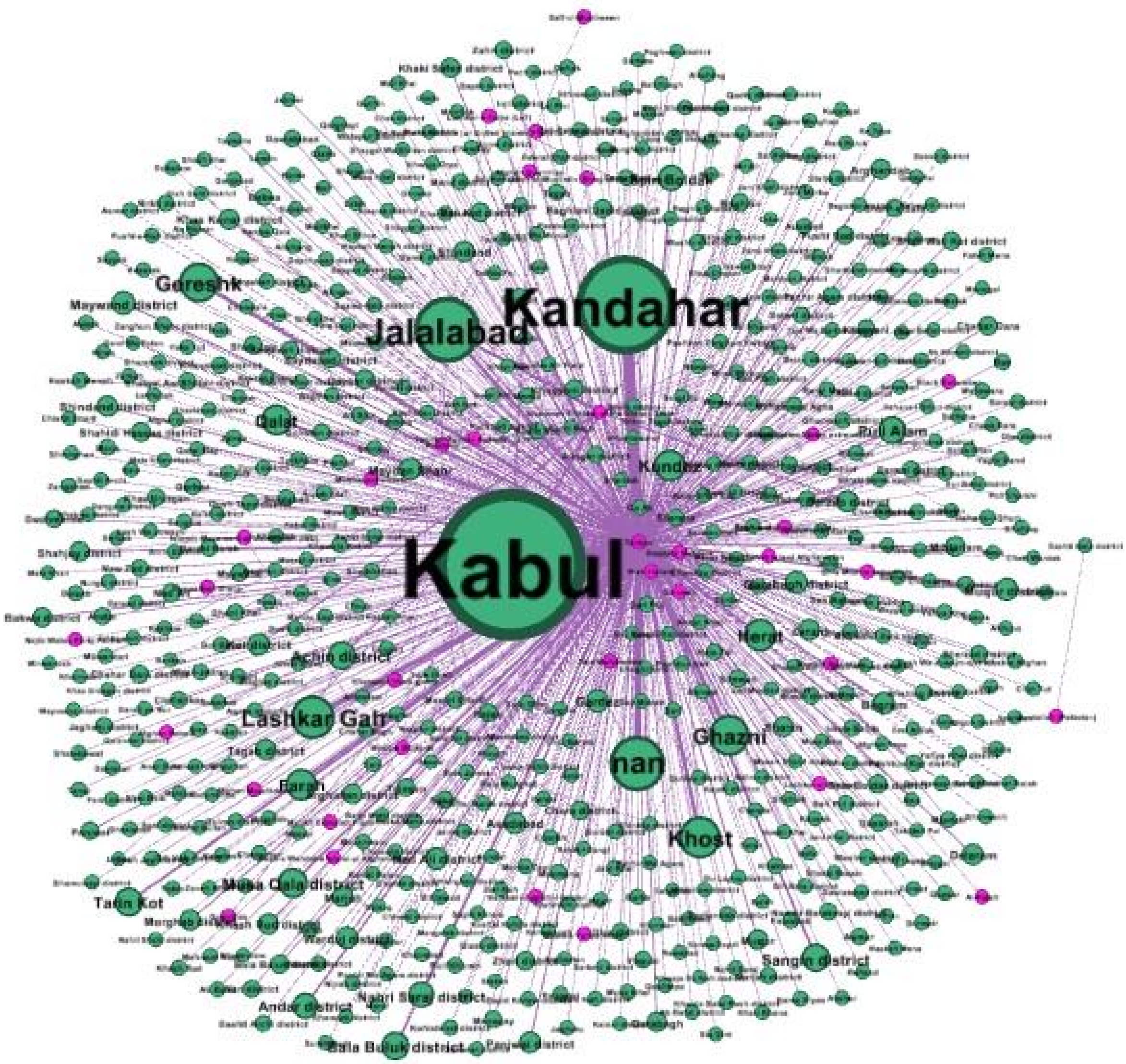
3.2 Afghanistan Terrorism Network

3.2.1 Overview

First we shall conduct a general analysis of Afghanistan with the following graph linking Terror groups to cities.

Figure 7

Afghanistan Terrorism Network: Overview



Note. This is a bimodal city-attackers network, the green nodes represent cities and the pink represents the attackers.

The graph above shows the overview of the Afghanistan terrorism network. In the middle of the network, you can see the Taliban organization as the center of the network. It has the highest degree and the weighted degree shows that it has a huge influence on the Afghanistan terrorism network.

Table 7

Afghanistan Terrorism Network Sorted by Weighted Degree

Id	Degree	Weighted Degree	type
Taliban	567	2596.0	gname
Kabul	22	249.0	city
Khorasan Chapter of the Islamic State	54	192.0	gname
Kandahar	5	154.0	city
Jalalabad	10	95.0	city

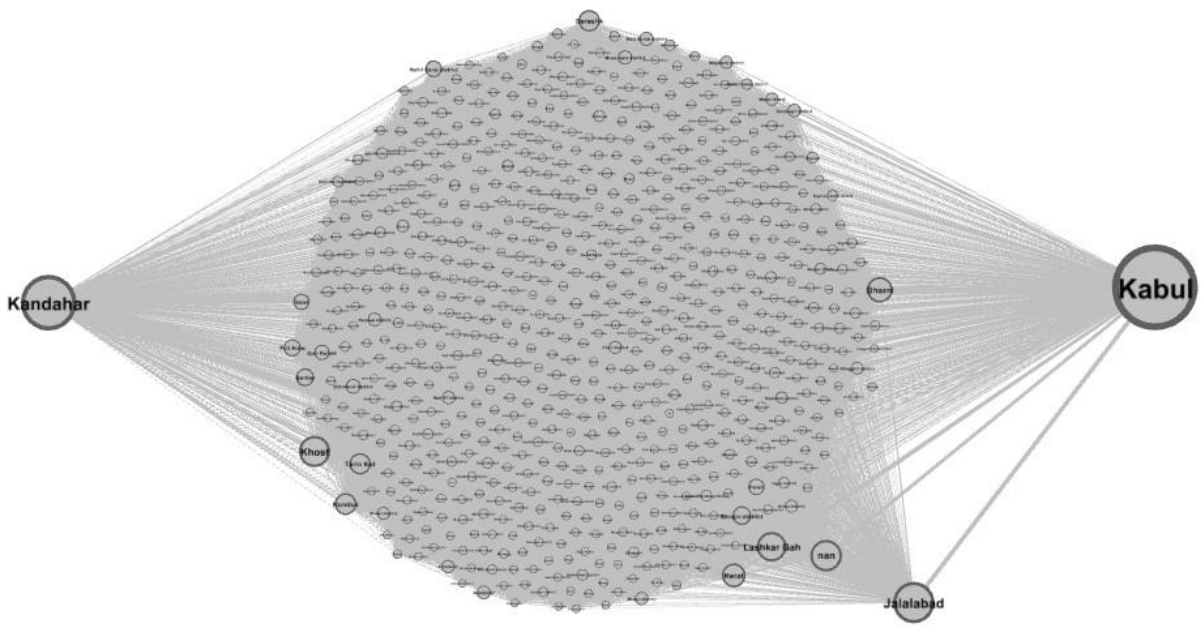
On the other hand, Kabul is the city with the most past terrorist attacks. It occurs in 22 different organizations (shown by its degree) and on 2596 different occasions. Since the same terror group is having a really high edge weight and number of attacks in Afghanistan, it shall be more useful to analyze relationships among cities. Thus, we conducted a city-city network graph which aimed to see which cities in Afghanistan are getting attacked by the same terror organization. Since terror organizations have political agendas and attack similar cities together, the graph shall be useful to visualize which cities are targeted more and together.

3.2.2 Terrorist Attacks in Afghanistan Cities

Now we shall conduct a more city specific analysis.

Figure 8

Terrorist Attacks in Afghanistan Cities: City-City Network



Note. The graph shows all the cities in Afghanistan that have been attacked by terrorist groups, and links between them show that they have been attacked by the same organization/terrorist groups. The size of the node in this graph is proportional to the degree.

From the city-city network shown in figure 8, we can see that Kabul and Kandahar are the most connected important cities in this city-city network. However, to understand further on the graph above, we conduct a further analysis below:

3.2.2(a) Graph Density Analysis

The graph density of this graph is very high i.e. 0.865, in comparison to our previous graph which represented general analysis of Terror groups across the globe. A much higher graph density means that the cities are more interconnected and the number of edges in the graph is close to the maximum number of edges. This shows that many cities are interconnected and thus all cities are being attacked together by the terror groups. This is highly concerning as the same cities are being attacked by almost all terror groups, instead of terror groups attacking cities specific to their political agendas.

3.2.2(b) Frequency and Degree Analysis

Table 8

Afghanistan Terrorist Attacks Data Sorted by Frequency

Id	Frequency	Degree	Weighted Degree	Closeness Centrality	Betweeness Centrality	Clustering Coefficient	Number of Triangles	Eigenvector Centrality
Kabul	462	708	829472.0	1.0	722.4553883731098	0.8674434029359352	217102	1.0
Kandahar	284	690	676631.0	0.9752066115702479	118.53362731904662	0.9104141688227004	216410	0.9982627842785107
Jalalabad	198	706	374594.0	0.9971830985915493	453.7696740873989	0.8723163160749804	217089	0.9999400092147518
nan	139	693	347614.0	0.979253112033195	387.9184758038992	0.9026933246586426	216446	0.9983950097131531
Khost	137	690	332813.0	0.9752066115702479	118.53362731904662	0.9104141688227004	216410	0.9982627842785107

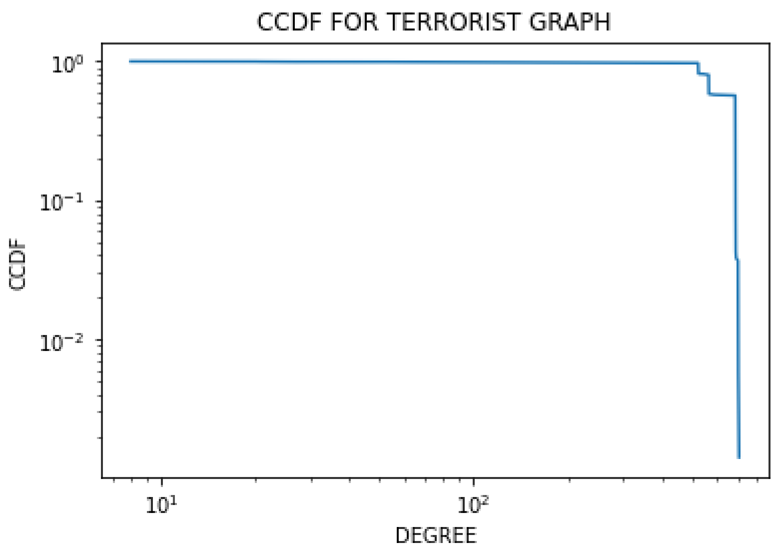
From the above tables it is observable that the Capital City 'Kabul' has the highest frequency as well as degree. This shows that it interconnected to the most number of cities across Afghanistan out of all cities in the network. Thus, we can also conclude that terror groups who attack any city across Afghanistan are also very likely to have attacked 'Kabul'. Kabul has also the highest weighted degree across all cities which again represents the same conclusions. In general the average degree of this network is very high and thus all the nodes are connected to a large number of other nodes meaning the same terror groups are attacking similar cities and these cities are thus connected to each other.

3.2.2(c) Degree Distribution for the Network

Many observations cannot be made from the CCDF of the network, which is as follows.

Figure 9

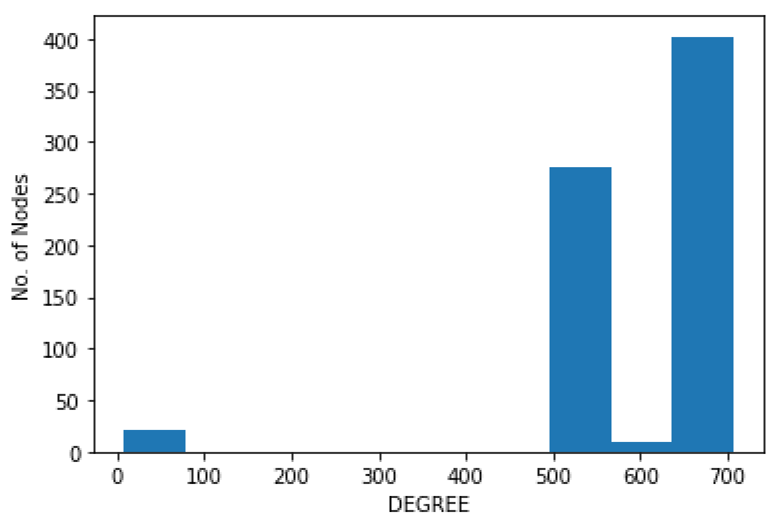
Terrorist Attacks in Afghanistan Cities: CCDF Distribution



As shown above, this graph is not very useful for conducting further analysis. Thus, it shall be better to plot a histogram showing degree distribution

Figure 10

Terrorist Attacks in Afghanistan Cities: Degree Distribution in histogram



Thus as we can see most nodes in this network have a degree in the range of 500-700. This refers to the fact that most cities are connected to 500-700 other cities which are attacked by the same terror group. This indicates that the number of edges (meaning degree) for each city is almost equal to the maximum degree that it can have. This reiterates the fact that the Afghanistan network is extremely dense with the degree for many nodes being close to maximum degree.

3.2.2(d) Centrality Analysis

Firstly considering Closeness Centrality, we view that Kabul has a closeness centrality of perfect 1 which refers to the fact that the shortest path of Kabul to all nodes is exactly 1. Kabul is thus connected to each and every node in the network, meaning any terror group which has attacked any city across Afghanistan has definitely also attacked Kabul, the capital city. This is an extremely worrying sign as it shows that each terror group which has carried out attacks across Afghanistan has definitely attacked Kabul. Other cities including Jalalabad and Herat also have really high closeness centrality and are interconnected with many cities.

Table 9

Afghanistan Terrorist Attacks Data Sorted by Closeness Centrality

Id	Frequency	Degree	Weighted Degree	Closeness Centrality	Betweeness Centrality	Clustering Coefficient	Number of Triangles	Eigenvector Centrality
Kabul	462	708	829472.0	1.0	722.4553883731098	0.8674434029359352	217102	1.0
Jalalabad	198	706	374594.0	0.9971830985915493	453.7696740873989	0.8723163160749804	217089	0.9999400092147518
Herat	93	704	218569.0	0.9943820224719101	376.8234122652866	0.8771135070477176	217047	0.9998074314072906
Khogyani district	27	704	62549.0	0.9943820224719101	373.6725980640086	0.877125630415104	217050	0.9998138079874503
Ghazni	104	702	260905.0	0.9915966386554622	298.39300290856124	0.8819634953729105	217008	0.9996812301799878

Figure 11

City-city Network: Classified based on Betweenness Centrality

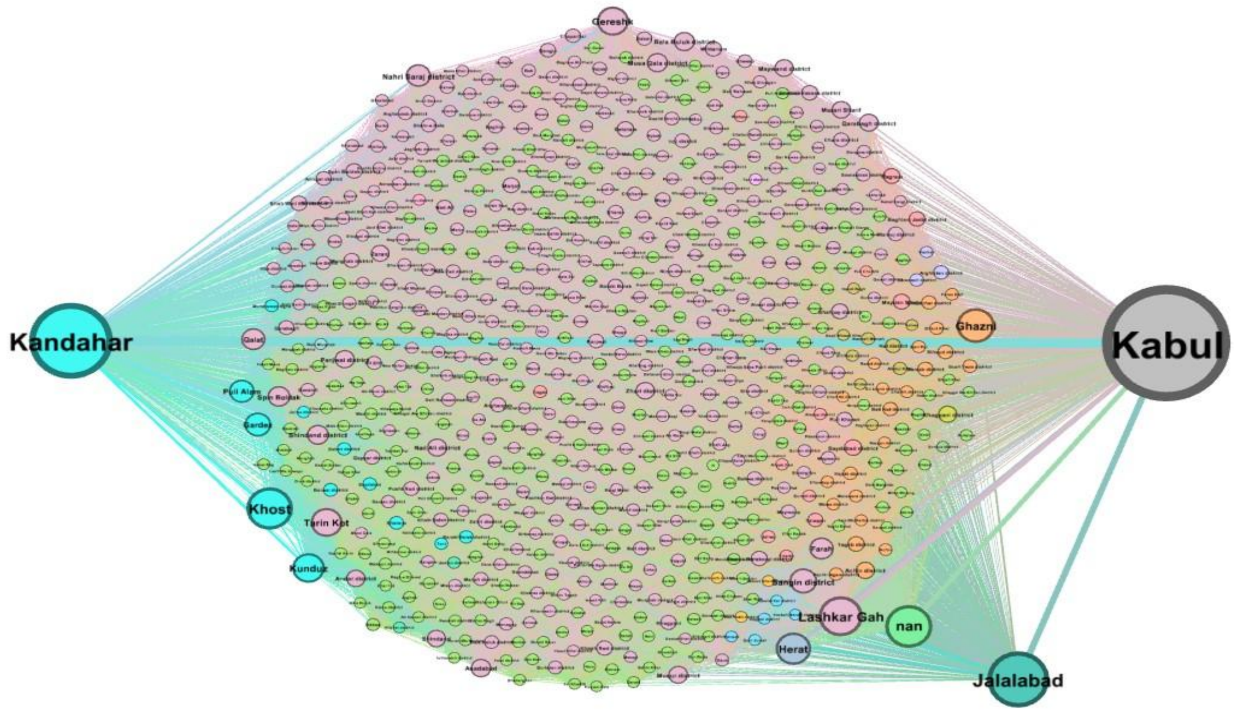
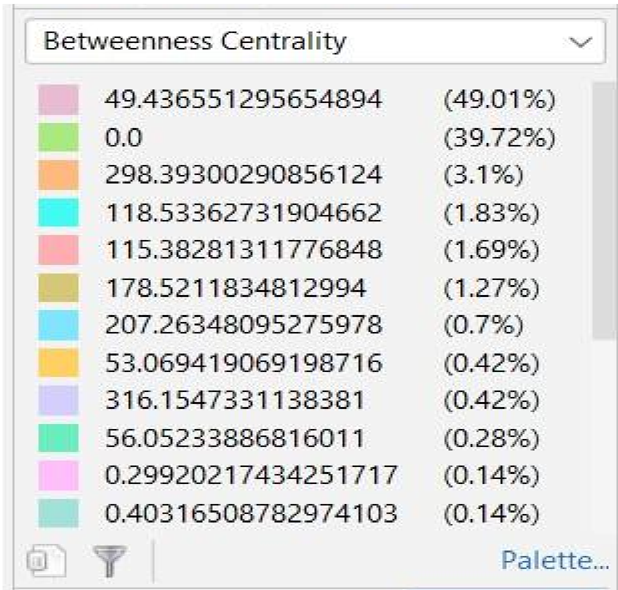


Figure 12

City-city Network: Classified based on Betweenness Centrality's legend



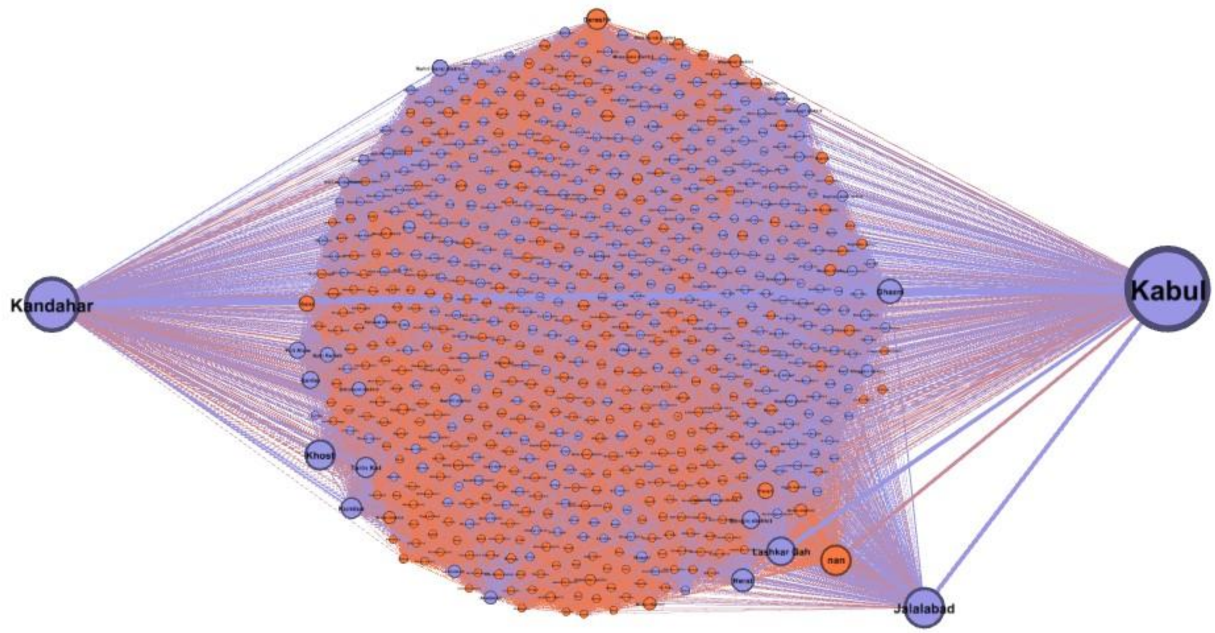
Similarly, when we look at Betweenness Centrality, Kabul and cities like Jalalabad and Herat have the highest Betweenness Centrality, thus these cities, because of having a very highest degree, highly contribute to the influence they have over the information flow through the network and act as bridges between different nodes.

The graph below describes how the graph can be divided according to betweenness centrality. Most nodes are in the dark pink color and as observable from the palette represent nodes having average betweenness centrality of 49.4. It is also interesting to see that green nodes have an average betweenness centrality of 0. There are also nodes in orange and light blue color which have really high betweenness centrality. Thus we conclude that most nodes of this network are having an average betweenness centrality which is not so high, the nodes having extremely high betweenness centrality only constitute a small portion of the total nodes.

3.2.2(e) Modularity Analysis

Figure 13

City-city Network: Classified based on Modularity



Note. This is a city-city network, the color represents the group based on modularity, if we look into the details we can see green nodes which include only 0.14% of nodes in the data.

It is interesting to see that the network has a low modularity score i.e. 0.05, the modularity score is used to measure the extent to which a network can be divided into different modules and clusters and also how well the nodes have dense connections within the cluster but sparse connections outside the cluster. In this network, the average degree of all nodes is very high and the network is very dense, thus the ability to further differentiate the network into different clusters is also low and even when nodes are clustered into two separate clusters as shown in the image, they constitute a lot of links to nodes in the external clusters as well. Thus the modularity of this network is quite low because of these reasons.

3.2.2(f) Why are the number of triangles so high?

Another interesting observation about this specific network is the extremely high number of triangles among the cities. The extremely high number of triangles basically states that the cities also do exist as a triplet and cases wherein the same 3 cities have been attacked by the same terror organization also occur often in this network. This again is correlated to how this network is extremely dense having a large number of triangles and a high density.

Table 10
Afghanistan Terrorist Attacks Data Sorted by Number of Triangle

Id	Frequency	Degree	Weighted Degree	Closeness Centrality	Betweeness Centrality	Clustering Coefficient	Number of Triangles	Eigenvector Centrality
Kabul	462	708	829472.0	1.0	722.4553883731098	0.8674434029359352	217102	1.0
Jalalabad	198	706	374594.0	0.9971830985915493	453.7696740873989	0.8723163160749804	217089	0.9999400092147518
Khogyani district	27	704	62549.0	0.9943820224719101	373.6725980640086	0.877125630415104	217050	0.9998138079874503
Herat	93	704	218569.0	0.9943820224719101	376.8234122652866	0.8771135070477176	217047	0.9998074314072906

4. RESULTS

4.1 Summary of results based on main country vs terror group network

- The graph is very sparse, meaning the density of the graph is very low, this tells that each of the terror groups attacks specific countries and not all based on their political agenda
- Certain nodes have very high degrees meaning these countries like France are attacked by a large number of different terror groups, even though the number of attacks may not be very high, meaning that for these nodes the weighted degree was not so high. Thus, such countries are being targeted by a large number of organizations with different political agendas and planning against terrorism requires specific strategic planning.
- There are certain nodes which do not have a high degree but their weighted degree is very high. These countries do not have many terror groups which target them, but the number of attacks by the terror groups is very high as in many cases these terror groups operate within that specific country and target individuals or groups within that country.
- On analysis of the degree distribution of this network, it can be seen that truncated power law fits the distribution the best; however, in comparison with exponential and power law the p value was not below 0.05, so the difference is not so significant.
- An eigenvector centrality analysis was also conducted which showed how most countries have a very low eigenvector centrality. An interesting observation was Lebanon having the highest eigenvector centrality, which refers to how important the node's neighbors are to the network. Thus even Lebanon's weighted degree is not so high meaning that it has not been attacked many times by terror groups, its high eigenvector centrality explains how it is being attacked by terror groups that contribute most to the social network.

4.2 Summary of Afghanistan Network Analysis

- First a general network graph was plotted showing terror groups and cities in Afghanistan, which showed how even though many terror groups are present in Afghanistan, it's Taliban having the highest influence, the highest weighted degree as well as the degree.
- Based on the correlation of cities analysis, it was found that the graph density is extremely high. This tells us that cities have edges close to the max number of edges. Thus, many cities are interconnected which represents how terror groups have not specifically attacked specific cities but attacked all cities in Afghanistan, this is highly in contrast to the Country vs Terror group attack wherein each terror group had attacked a specific group of countries.

- An interesting observation was that the betweenness centrality of the capital city, Kabul, was perfectly 1, which represented how Kabul was connected to each and every city in this network. This is an important conclusion as it shows that all terror groups which have attacked Taliban over the past years have definitely attacked the capital city Kabul.
- The modularity analysis showed the modularity score of the network to be very low which displayed that grouping the network into subnetworks is tough due to high interconnectedness among nodes of the graph, this displayed that for governments planning against terror requires a common strategy across the country since the country cannot be broken down into clusters which have common properties within themselves.
- On Degree distribution analysis, CCDF of the graph in Log - Log scale was unable to provide us much information, so it was more useful to plot a histogram showing degree distributions which also shows and reiterates the fact that average degree of the network is very high and most nodes have degrees in the range of 500-700.
- The number of triangles in the network was also very high which represented and reiterated the fact that the network density was very high in the network and cases wherein 3 nodes are interconnected is a common feature of the network.

5. Conclusion

Through this project we have described our analysis of worldwide terror attacks by different terror groups as well as an in depth focus on Afghanistan, which is the country most affected by terrorism in this world.

Through our analysis, we reached important and key conclusions which can help in our fight against terrorism. These insights can help in strategic planning by governments and organizations across the globe. Firstly our global analysis shows how different terror groups have had specific political agendas and focussed on specific countries and not attacking every country. In some cases countries like France had really high degrees which referred to how they had been targeted by a lot of different terror groups, although the weighted degree was not so high meaning even though many terror groups had attacked, the number of attacks by each terror group was not so high. In some other cases the weighted degree was very high for countries but the degree was low, which referred to how these countries had been attacked by a few terror groups but the number of attacks was very high. Both of the above types of countries require very different kinds of attention, countries like France are attacked by a lot of organizations because of Historical reasons, and in order to further understand the second group which included countries like Afghanistan, our team did a country specific analysis. There were also cases like that of Lebanon, which did have the highest weighted degree but the centrality and referred to how even though the number of attacks on Lebanon were not very high they were attacked by organizations who contribute most to the network.

Now moving to our second case of Afghanistan and Taliban, it is important to understand that the Afghanistan-Taliban conflict involves a complex history. The Taliban is a fundamentalist Islamic movement that emerged in Afghanistan whose ideology is based on a strict interpretation of Islamic law and a desire to establish a pure Islamic state (Mullah, 2019). The Taliban became a major terrorist group in Afghanistan after taking control of the country in 1996. There are many other terror groups operating in Afghanistan including Al Qaeda. In order to conduct a more specific analysis, and an analysis which could show correlations between cities in Afghanistan, the team conducted an analysis on city - city social networks in Afghanistan.

Based on our data analysis, Kabul and Kandahar are the two cities most frequently targeted by terrorist attacks in Afghanistan. There are several reasons why these cities are particularly vulnerable. Firstly, both cities are densely populated, making it easier for terrorists to blend in and carry out attacks. Secondly, both cities are home to key government institutions and infrastructure, including military bases and government buildings. Terrorist groups target these institutions as a way to undermine the government's authority and create chaos (Shams, 2018). Finally, both cities are located in regions with a significant presence of terrorist groups. The

Taliban, in particular, has a strong presence in both cities and has carried out numerous attacks in the past.

Also interesting conclusions including how every terror group which has attacked any city in Afghanistan has also attacked Kabul, can be used by governments to understand the political motives behind many terror groups. Our analysis shows how clustering the Afghanistan network into subnetworks is not easy and thus the whole country requires a similar approach in dealing with terrorism as the similarity among cities in terms of being attacked by the same terror group is high. Usually government organizations have different approaches for different localities in the country based on different terror groups operating in those areas. However, the case is different for Afghanistan, with it being an extremely dense network, terror groups seem to have targeted all cities instead of focusing on specific ones.

Finally, the team also conducted predictive analysis based on link prediction utilizing Node2Vec algorithm to embed the nodes of the graph. The whole project in conclusion was aimed to foster a better understanding of terrorism landscape across the globe while also focusing on Afghanistan. It is hoped that this project can help organizations across the globe develop a better understanding and also prove as an aid while countries participate in strategic planning in their fight against terrorism.

REFERENCES

BBC News. (2021, September 28). *Why is there war in Afghanistan?* https://www.bbc.com/news/world-asia-20232853

Council on Foreign Relations. (2022, February 23). The Afghanistan War. https://www.cfr.org/backgrounder/afghanistan-war

Encyclopedia Britannica. (n.d.). *War in Afghanistan* (2001–2021). https://www.britannica.com/event/war-in-Afghanistan-2001-2021

Global terrorism database. (n.d.). Kaggle: Your Machine Learning and Data Science Community. https://www.kaggle.com/datasets/START-UMD/gtd

GlobalSecurity.org. (n.d.). *The US War in Afghanistan.* https://www.globalsecurity.org/military/ops/enduring-freedom.htm

Mullah, Z. (2019). *Taliban in Afghanistan: History, Ideology and Objectives*. Journal of Political Studies, 26(2), 49-67.

Shams, S. (2018). *The War in Afghanistan: The Taliban and the Afghan Government*. Middle East Policy, 25(2), 123-135.