Fight against terrorism

Network Tour of Data Science

Team 54:

Khalil Elleuch
Mariem Belhaj Ali
Mouadh Hamdi
Nour Ghalia Abassi

Introduction

- Marked increasing acts of terrorism across the globe in recent decades..
- Significant number of victims due to this acts.

What are the relationships between the attacks?

Can we find significant communities sharing specific properties?

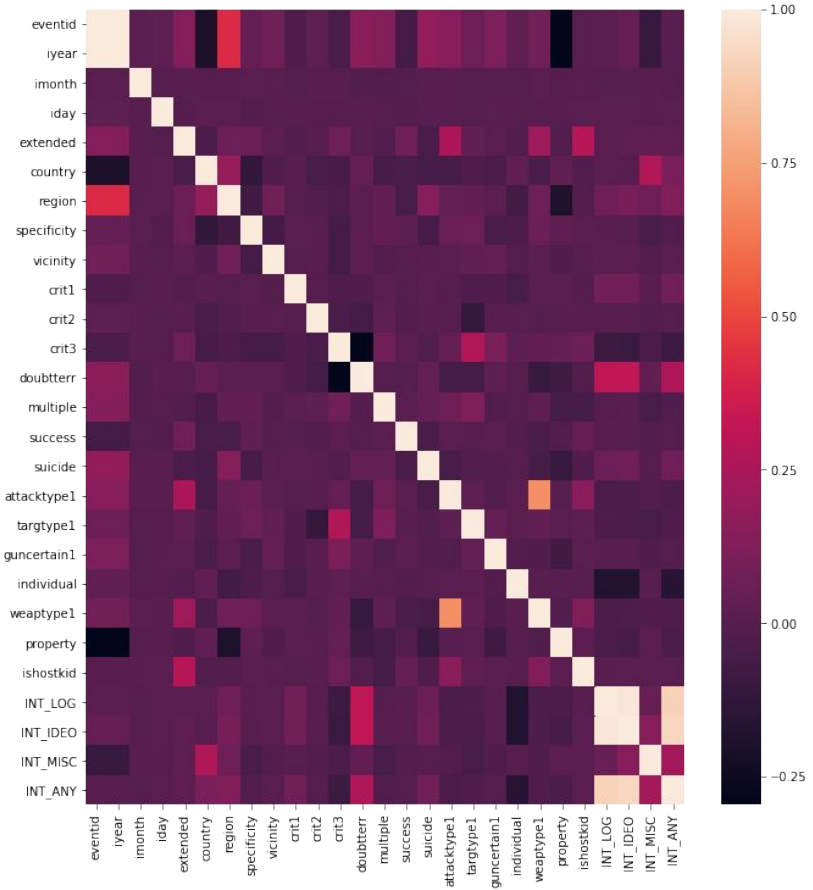
Dataset

Global Terrorism Database

- Open-source database including information on terrorist attacks.
- The GTD includes data on domestic as well as international.
- More than **180,000** attacks.
- **1970** ⟶ **2017**.

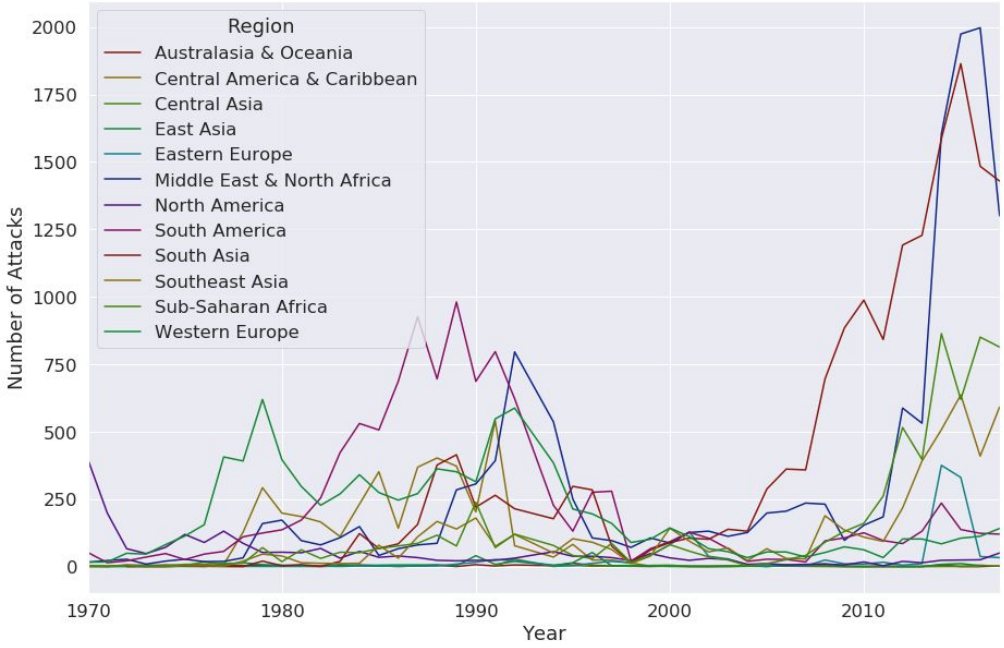
Data Cleaning

- Delete 45% of data containing ”Unknown”s in Group Name (groups involved in the terrorist attacks).
  - They have a huge proportion in the dataset.
  - We will use those values to make predictions.
- Studying correlations between features.



Data visualisation

- The number of the terrorism attacks get increased exponentially in the past 20 years
- All continents/ countries are concerned!

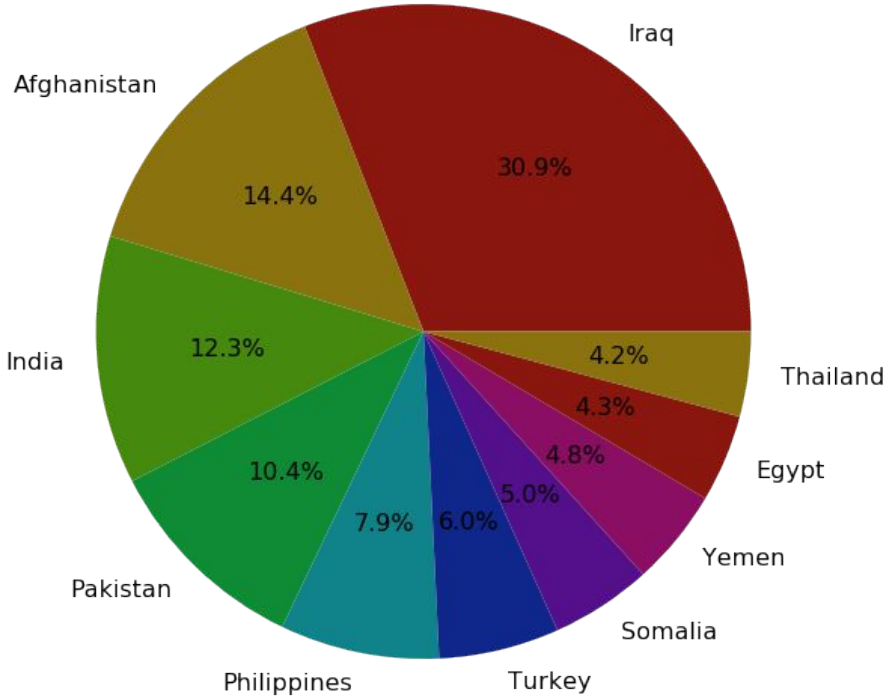


Evolution of the attacks per region

Data visualisation

Still some countries are more concerned than others:

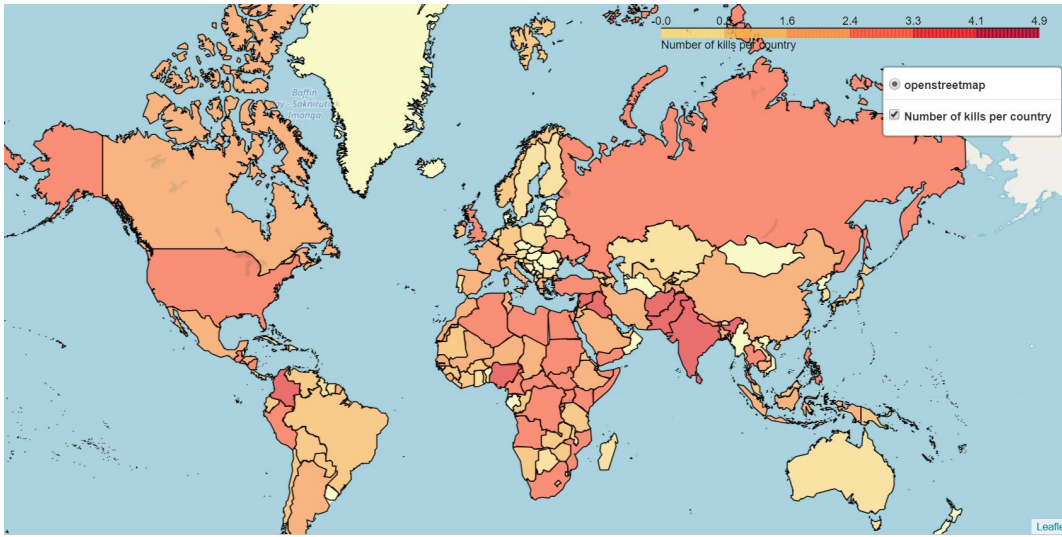
- Iraq is leading with almost ⅓ of terrorist attacks!
- Developed countries are the less present in our dataset.



Percentage of attacks in 2016 in each country

Victims:

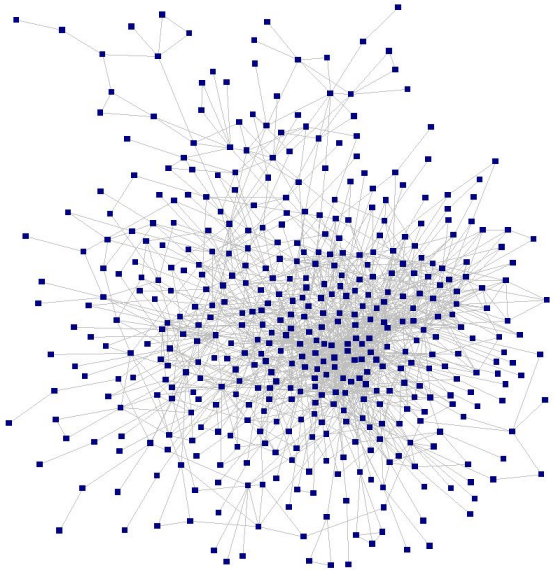
- Middle East: A lot of wars during the last century.
- Africa: This can be explained by the huge number of civil wars.
- USA/ Russia



Number of kills per country

Network analysis

- Cluster terrorism attacks according to the targets
- Check if a group always attacks the same targets.
- Capture if there is a collaboration between terrorist group.



Network analysis

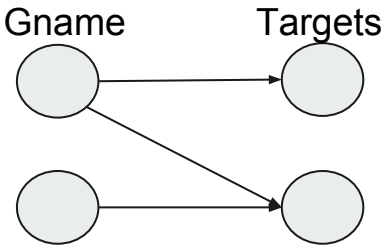
Targets: The victims of the attacks

targtype1_txt	Nb attacks
Private Citizens & Property	43511
Military	27984
Police	24506
Government (General)	21283
Business	20669

Organization name: Terrorist organization

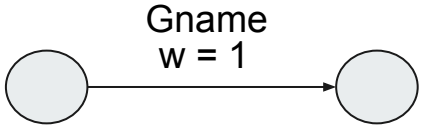
gname	Nb attacks
Taliban	7478
Islamic State of Iraq and the Levant	5613
Shining Path	4555
Farabundo Marti National Liberation Front	3351
Al-Shabaab	3288

Graph construction



1. Create a bipartite graph

- **Nodes:** The organizations names given in gname feature and the targets.
- **Edges:** Connects each organization with its target.



2. Weighted projection to gname

- **Nodes:** The organizations names.
- **Edges:** Connects two organizations if they attacked the same target.
- **Weights:** Number of times the two organizations attacked the same target.

Network properties

Nodes importance:

Page Rank: Computes a ranking of the nodes in the graph based on the structure of the incoming and outgoing links.

Degree Centrality: Assumes that the node with the most connections (edges) is the most important.

Network properties

Uncover important key players in the graph with nodes importance properties:

- The same top 10 terrorist organizations appear to have highest page rank and highest degree centrality in our graph.

- Some world known terrorist organization such as Kurdistan Workers' Party, Taliban and Hezbollah seems to have an important role in terrorist attacks.

Kurdistan Workers' Party
Tehrik-i-Taliban Pakistan
Basque Fatherland and Freedom (ETA)
Liberation Tigers of Tamil Eelam (LTTE)
Taliban
Maoists
Al-Shabaab
New People's Army (NPA)
Palestinians
Hezbollah

Clustering

Louvain community detection

- Method to extract communities
  from large networks (greedy optimization) .
- Unfortunately, we didn't get anything relevant using this algorithm.
  → no real separation between nodes and no clusters were detected.

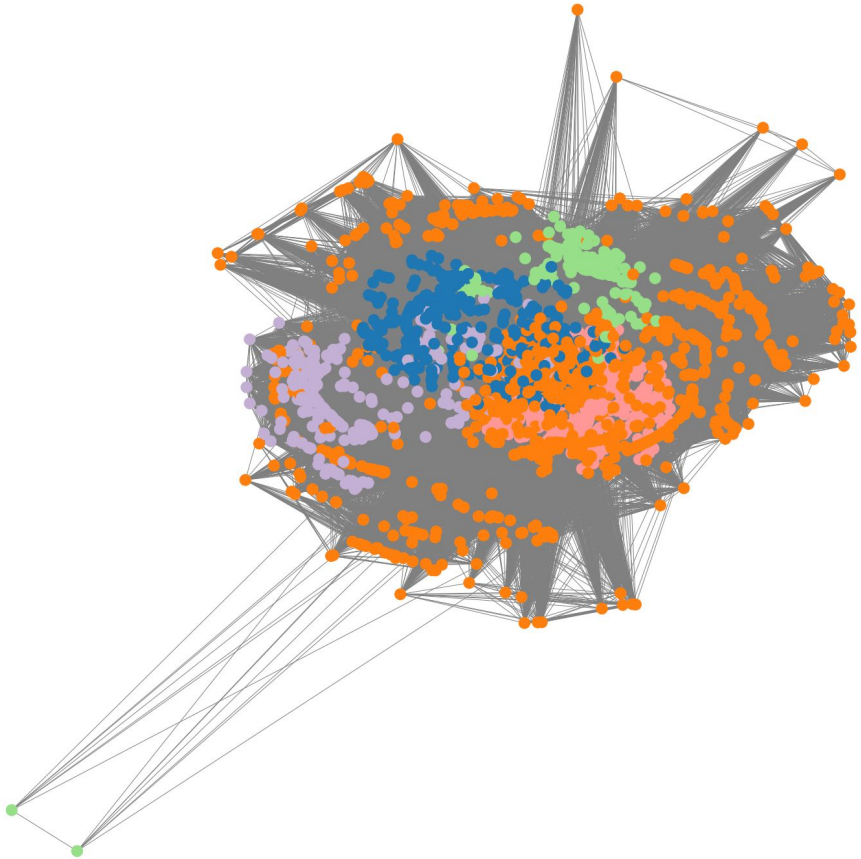
Clustering

Spectral clustering

- Use of the **spectrum** (eigenvalues) of the **similarity matrix** of the data to perform **dimensionality reduction** before clustering in fewer dimensions.

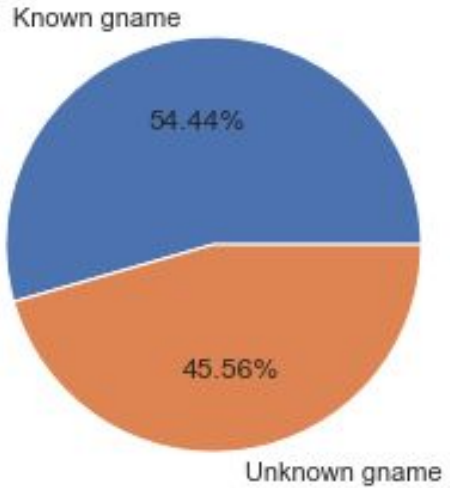
- In our case, we get an interesting result :

  5 clusters found using spectral clustering.



Machine Learning

- Predict the 'Unknown' groups considered as our missing data.
- Select the 15 most important gnames and assign the remaining to a new class 'other'.
- Number of unknown data: 82385.



⟶ 0.62 Accuracy to predict the unknown gnames.

Machine Learning

Predict the unknown group names using the spectral clusters

- Assign each “unknown” to a cluster not a group name.
- Split data set of known gnames into train/test set.
- Train a random forest classifier using the cluster given by the spectral clustering.

⟶ **0.88** Accuracy to assign each unknown gname to a cluster of organisations names

Conclusion

- Importance of network analysis is to have insights on aspects of the data that we couldn't investigate.
- Machine learning procedures helps predict unknown terrorist organizations.
- Terrorism is attacking every country in the world : Governments need to rise awareness and invest more in fights against terrorism.

Thank you for your attention !

Any questions ?

You can find the link here :
https://github.com/mouadhhamdi/NTDS_Project?fbclid=IwAR0NDYzjE5HvWSpWamiUtZVQhb3I7IbDTb5hIdqUdOlDwfrdPje1RGpTY-A