Predicting Terror Attacks?
A Network Story

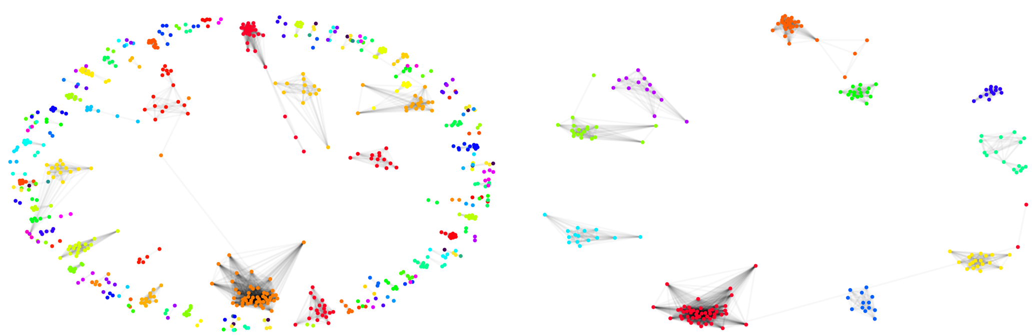
Enea Figini, Axel Nilsson, Nicolas Bollier, Elias Le Boudec
Team 29

January 17, 2019

1 Introduction

2 Exploring the Data

2.1 Relationships Dataset



(a) *Terror attacks location graph, colouring by component ID*

(b) *Ten biggest components from the terror attacks location graph*

Figure 1: *Graphs analysed in the project*

2.2 Terror Attacks Dataset

The formation of the network implies a transitive relation between most of the nodes. Indeed, if for most nodes, take a b and c in the network, we have

$$a \sim b \text{ and } b \sim c \text{ then } a \sim c \tag{1}$$

Equivalently, if attack a took place close to b, and attack b took place close to c, then it is probable that attack a took place close to c.

3 Data Quality

3.1 Terror Attacks Dataset

Multiple issues regarding data quality have been found in this dataset:

Broadness The dataset comprises attacks ranging from 1969 to 1950 and spanning the entire globe. Simple and relevant explanations for the graph formation or properties are not likely to be found, since the mechanisms behind two different attacks can be entirely different.

Structure Half of the nodes are isolated, hence the topological information they carry in the graph is very limited. What is more, because of the transitivity relation described in Section 2.2, connected components are in most of the cases complete, hence isotropic.

Reliability Errors have been found in the data. For example nodes `Djibouti_-Youth_Movement_19900927` and `Armed_Islamic_Group_19950711` have been connected, whereas the first attack took place in Djibouti [1] and the second one in Paris [2]. Hence algorithms using the data must tolerate some error in order to avoid overfitting.

4 Predictions

The algorithm used to predict the terror attack location is the following:
Let w be the application that returns a weight for each pair of nodes (n_1, n_2) in the graph $\mathcal{G}$, defined as

$$w : \mathcal{G}^2 \to \mathbb{R}^+ \tag{2}$$

$$(n_1, n_2) \mapsto f(|n_1 - n_2|) \tag{3}$$

where

$$|n_1 - n_2| = \|\text{features}(n_1) - \text{features}(n_2)\|_2 \tag{4}$$
$$\text{features}(n) \text{ is a binary features vector for each node } n \text{ in } \mathcal{G} \tag{5}$$
$$f:\ \mathbb{R}^+ \mapsto \mathbb{R}^+ \text{ is a decreasing function} \tag{6}$$

Examples for f are given in Table 1.

Algorithm 1: *Finding the predicted location of the next terror attack*

Data: Set of connected components $\{C_i^t\}$, $i = 1, \ldots, 10$, and the features vector of the next terror attack n_{t+1}, i.e. feature(n_{t+1}), at each timestep t

Result: Location prediction p_t for time $t+1$ at time t, at each timestep t

for *each timestep* t **do**

  Compute the lead component $l(C_i^t)$ for each component C_i^t

  $p_t = \arg\max_{i=1,\ldots,10} w(n_{t+1}, l(C_i^t))$

end

return p_t

Algorithm 2: *Finding the lead node of a connected component with weighted edges*

Data: Connected component C

Result: Lead node n_l

Initialise $s(n)$ to zero. s is a dictionary mapping a score $s(n)$ for each node n

for *each edge* e *from* C **do**

  Let $e = (n_1, n_2)$, w be the weight of e

  $s(n_1) \leftarrow s(n_1) + w$

  $s(n_2) \leftarrow s(n_2) + w$

end

return $n_l = \arg\max_{n \in C} s(n)$

Table 1: *Prediction accuracy for different node distance weightings*

Weighting		**Best skewness ζ**	**Accuracy**
Gaussian:	$w = e^{-d^2/\zeta} - e^{-1/\zeta}$	0.01	50.5 %
Log-Exponential:	$w = e^{-d} \log\left(\frac{1+\zeta}{d+\zeta}\right)$	0.1	50 %
Linear:	$w = 1 - d$	N.A.	47 %
Square:	$w = \begin{cases} 1 & d < \zeta \\ 0 & \text{otherwise} \end{cases}$	0.1	43 %

5 Conclusion

References

[1] Amnesty International Publications, 1 Easton Street, London, *Amnesty International Report 1991*, 1991.

[2] L'Obs, "Attentats de 1995 : chronologie." [fr] Online. https://bit.ly/2ASwNQP, last checked 17 January 2019, October 2007.