Predicting Terror Attacks? A Network Story

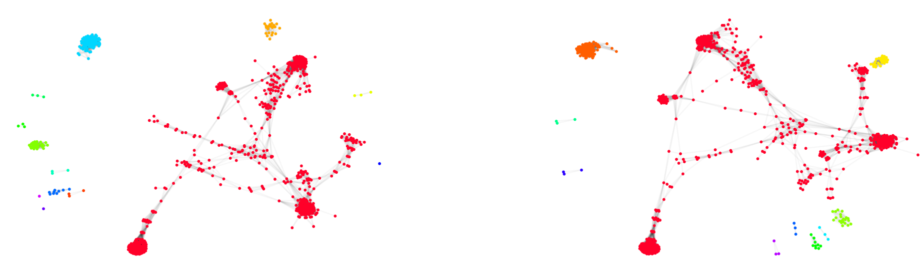
Enea Figini, Axel Nilsson, Nicolas Bollier, Elias Le Boudec
Team 29

January 17, 2019

1 Introduction

2 Exploring the Data

2.1 Relationships Dataset

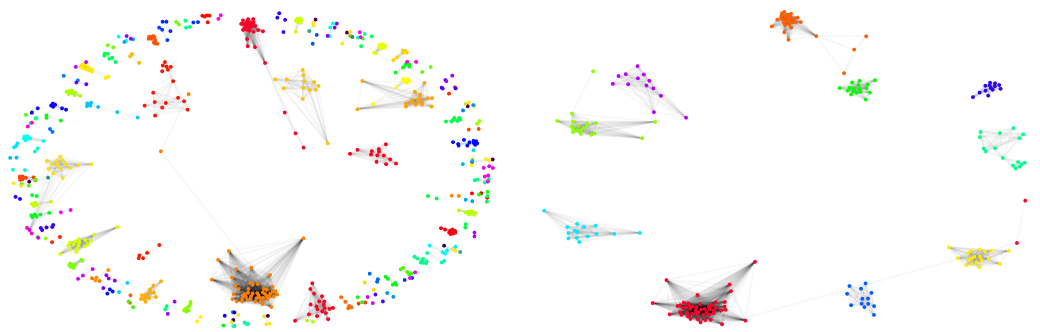


(a) *Terrorist relations graph, colouring by component ID*

(b) *Ten biggest components from the terrorist relation graph*

Figure 1: *Graphs of the terrorist relations dataset*

2.2 Terror Attacks Dataset



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

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

Figure 2: *Graphs analysed in the project*

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 [**?**] and the second one in Paris [**?**]. 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:
L

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