HY484 Project: Terrorist Networks

Visualization of networks created by attacks by the year.

Email: csd3805@csd.uoc.gr

Date: 15/01/20

1. ***Abstract:*** *We want to analyze data crafted by terrorist attacks (we will call them events) that took placed in various cities around the world by the year, so that we can identify the existence of terrorist networks. We will also be able to extract some conclusions about the properties of those networks. For reasons of imputation we are able only to analyze this huge dataset only by splitting in by the years. We will follow a specific approach to analyze and visualize our data.*

2. ***Our Approach:***

2.1. Data Analysis: *For analyzing the data we have created a python script in which we crunch the data following the next steps:*

2.1.1. *Firstly we have to analyze the format of the file where our dataset is stored .Here we had a csv file that holds information about the attacks. We apply some specific analysis using some conditions to hold data for a specified year. The data we want to analyze here are the succeeded events.*

2.1.2. *We extract information about an event and store them in a file named* **year_Attack.csv** *where year is the specific year we want to examine. The information we store is the following:*

a) event_id: *A unique id that characterizes one event*

b) city_id: *A unique id that characterizes a specific city*

c) gname: *The group name that claimed responsibility for the event.*

2.1.3. *We extract information about a city where an event has occurred in a file named* **year_City.csv** *where year is the specific year we want to examine. The information we store is the following:*

a) city_name: *The name of the city the event occurred.*

b) city_id : *A unique id that characterizes a specific city*

2.1.4. *We extract information about the connection between events based on the group that claimed responsibility, if two events were claimed by the same group then there is a connection between them. We store this information in a file named* **yearAttack_Attack_Connection** *where we hold the following data:*

a) group_id : This id has no connection with the group that carried out the event it is only a unique identifier for the connection.

b) from_attack: *An id to a specific event (source event)*

c) to_attack : *An id to a specific event (destination evet).*

2.1.5. *Finally we extract information for the connection between a city and an event. If an event let us say E has occurred to a city let us say C then there is a connection between E and C. The file that contains this information is named* ***yearCity_Attack_Connection***

a) city_attack_id: *An unique identifier that characterizes a connection between a city and an event.*

b) from_attack*: The event's id.*

c) to_city: *The city's id.*

2.2. Data Visualization: *Having crafted the four csv files we combine them as four sheets in an excel file named* ***year_DSAttacks.*** *This file will be used as our integrator for* ***Tom Sawyer Perspectives Designer.***

2.2.1. ***Integrator****: The excel file we have already defined.*

2.2.2. ***Schema****:*

2.2.2.1. ***Nodes:*** *The nodes in our graph will be the events that occurred in the specific year and the cities where events occurred for the same year. So we define 2 nodes :*

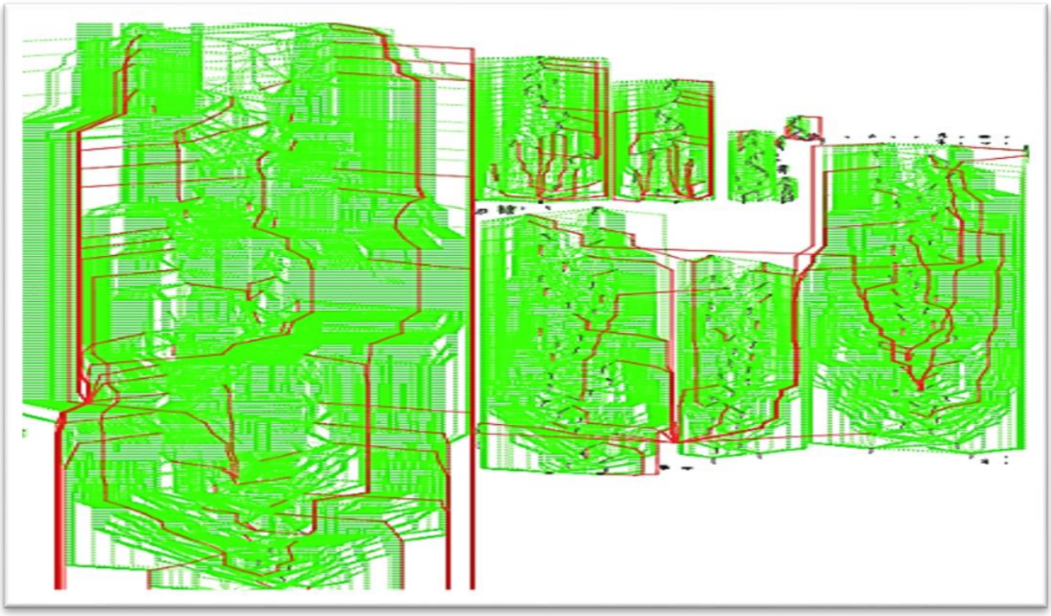
a) Attack: *sheet created from year_Attack.csv*

b) City: *sheet created from year_City.csv*

2.2.2.2. ***Edges:*** *In our graph there will be two kinds of edges. There are edges representing a connection between events and edges representing connections between events and cities:*

a)Attack_by_Group: *sheet created from* ***yearAttack_Attack_Connection.csv .***

b)Attack_by_City: *sheet created from* ***yearAttack_City_Connection.csv .***



3. Results: *Having created our integrators and schema we use the Tom Sawyer Prespective Designer to visualize our network. The following two figures depict the hierarchical layout and the circular layout of our network:*

Figure 1: Hierarchical Layout of our network

Figure 2:  ***Circular Layout of our network***

Firstly let us explain what we see here. The edges in our graph:

- ***a)*** *The green edges depict the connection between two events, meaning that two events are connected by a green edge is they were carried out / claimed by the same group.*
- ***b)*** *The red edges depict the connection between a city and an event, meaning that an event is connected to a city by a red edge if for the specific year this event was carried out in the specific city.*

The nodes in our graph:

- ***a)*** *Nodes that represent an event are depicted with an image of two swords and have information about the group that carried out / claimed the event, as depicted in Figure 3.*
- ***b)*** *Nodes that represent a city are depicted with an image of building and have information about city name where an event occurred, as depicted in Figure 4.*

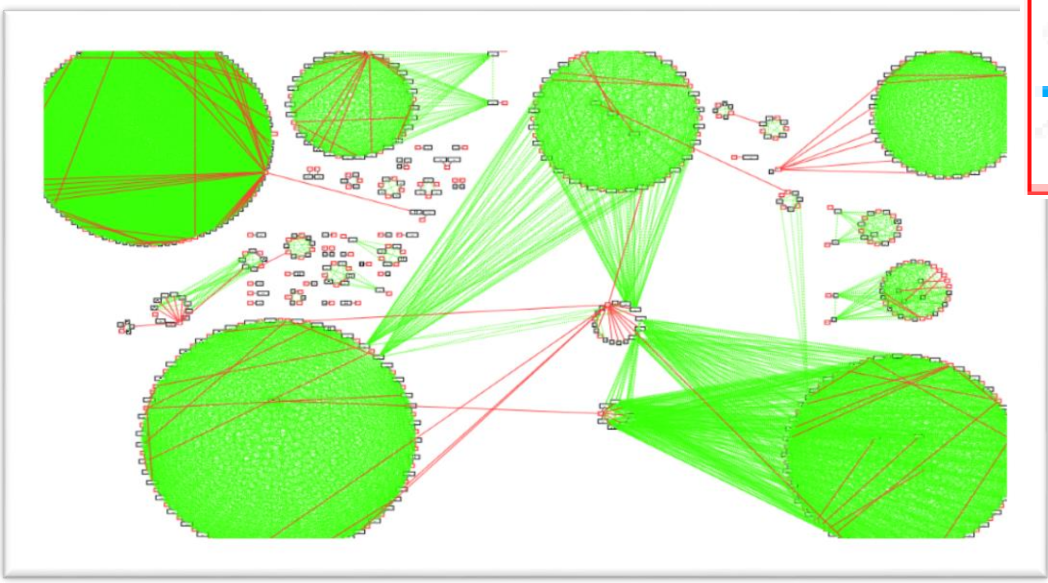


Figure 3: Node representing

Figure 4: Node representing

an event. ***a city.***

Here we can observe that the Hierarchical Layout (Figure 1) is not useful to determine any smaller-networks and draw conclusions .But after observing the Circular Layout (Figure 2) we are able to "see" clearly the existence of smaller networks that are represented as circles.

4. ***<u>Observations:</u>*** *By observing the circular layout of our network we can clearly see that there are 3 sets of connections:*
   - **4.1.** ***"Irrelevant" Events:*** *We call those events that have no green edges ''irrelevant'' as they do not depict the existence of a network .Those events are single events that have no connections with a group or an organization.*
   - **4.2.** ***''Random" Events:*** *Events that occurred in the same city but from different groups or organizations will be called random as they do not depicted a coordinated plan of events from a specific group to various targets.*
   - **4.3.** ***<u>Organized Cycles:</u>*** *This might be the most interesting observation of our graph analysis. Events that were carried out by the same group in a variety of cities-targets form a cycle inside our network that is a smaller complete graph. Those kinds of cycles have 3 interesting characteristics:*
     - **4.3.1.** *They depict a coordinated plan of attacks.*
     - **4.3.2.** *They are strongly connected and all the triangles they form are balanced.*
     - **4.3.3.** ***The density of the green edges depict the correspondence in that specific group. If there are fewer green edges in the graph that means that after a period of time that group was eliminated or the response to its threats were assessed faster and the attack was unsuccessful or stopped.***

It is also interesting to observe the 2016 graph (Figure 6) where although there is no significant difference in the number of cycles formed in the network with 2001 graph (2001: ~=22 cycles and 2016~=21 cycles) there is clearly a significant difference in the density of the coordinated events (green edges).

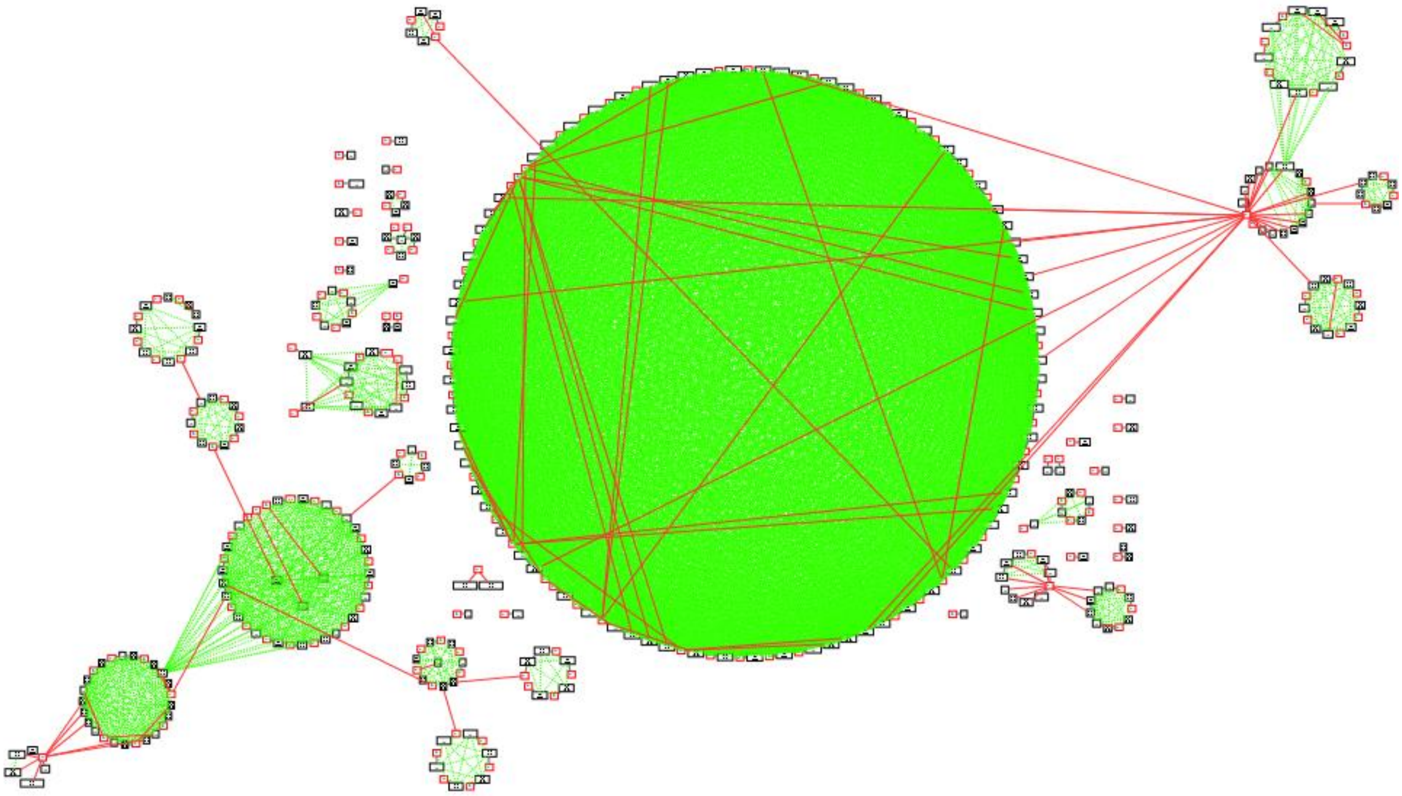


Figure 6: 2016 Graph

5. ***<u>Conclusions:</u>*** *We can fairly conclude to the following:*
   - **5.1.** *The existence of a cycle around a geographic location (a country) is most likely the result of its domestic or foreign policy. For example in 2016 Graph there is a huge cycle with many green edges that represent a revolutionary movement's actions against its government. Also a cycle around a specific geolocation can be formed is a country has an aggressive foreign policy against the country that roofs a group / organization.*
   - **5.2.** *The density of the green edges in a cycle is respective to the fact that if a group was eliminated then the according graph will include fewer green edges.*

6. ***<u>Useful links and Datasets:</u>*** *The dataset we used was* https://www.kaggle.com/START-UMD/gtd