Exposing the True Terrorist Network

NTDS Final Project

23.01.2019

Team 21: Timur Lavrov
Ayberk Tarçın
Sinan Yılmaz
Aslı Yörüsün

Motivation

We selected the Terrorist Relations dataset because:

- This type of data is scarce
- Importance of fighting terrorism
- Opportunity to perform social network analysis

Objectives

Objectives

- Examine how terrorists are related
- Predict unobserved terrorist relations
- Determine how relations affect the spread of terrorist ideas
- Identify organizational structures within the terrorist network
- Pinpoint the network's most wanted terrorist

Data Exploration

The original data represents the network as a line graph.

Nodes: Relationships (851)
Edges: Terrorists (8196)

Features: 0/1 binary values (1224)
Labels : Family(16%),
Colleague(54.2%),
Congregate(12.4%),
Contact(17.4%)

Pre-processing

Dual Graph

Pre-processing

- Nodes are represented with non-existing URLs. We parsed these URLs in order to extract the terrorists' names.
  - Some of the terrorists are represented as timestamps

    http://profilesinterror.mindswap.org/document/2005/09/28/23:42:30
    http://profilesinterror.mindswap.org/document/34#HassanNasrallah — Colleague
- The relationships are characterized by 1224 features
  - 612 distinct features assigned per terrorists
  - No descriptive explanations of these features

Dual Graph

In addition to the original graph, we created a *Dual Graph* which includes,

- individual terrorists as nodes
- relationships as edges.

851 nodes
8196 edges

→

244 nodes
851 edges

Constructing the Dual Graph we found 11 multi-labelled relationships as we obtained 840 edges between the 244 terrorists in the network.

Experiments

- *Revealing a greater network*
- *Identifying the top threat*

Expanding the observed network

As terrorists are secretive by nature and the provided dataset was purely constructed on empirical evidences, we tasked ourselves with predicting unobserved relations to enrich the network by:

- Reducing the feature set to 440 discting features, eliminating useless attributes
- Linking pairs of terrorists that have a positively correlated set of features

Original Dual Graph

244 nodes (terrorists)
840 edges (relations)

→

+ 105 edges
(12.5%)

Expanded Dual Graph

244 nodes (terrorists)
945 edges (relations)

Learning the labels of the predicted relations

Having expanded our Dual Graph we converted this back into a Line Graph representation to get the following enriched network:

Original Line Graph

840 nodes (relations)
8196 edges (terrorists)

Expanded Line Graph

945 nodes (relations)
10027 edges (terrorists)

We now had the grounds set to perform transductive learning in order to obtain the labels of the new relations. Taking a one-vs-rest approach for each label we obtained the following:

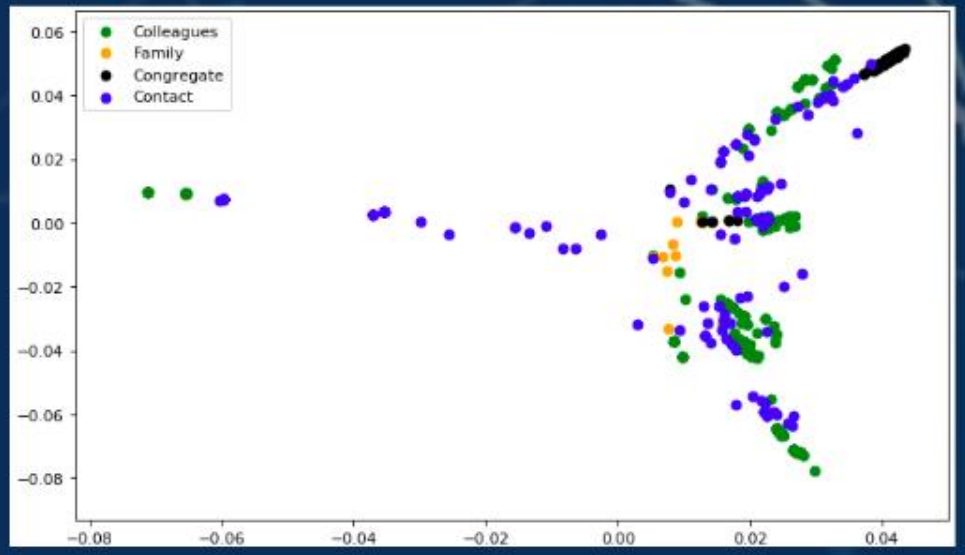
Label	Original	Expanded	% increase
Colleague	461	513	11%
Family	136	138	1%
Congregate	106	144	35%
Contact	148	157	6%

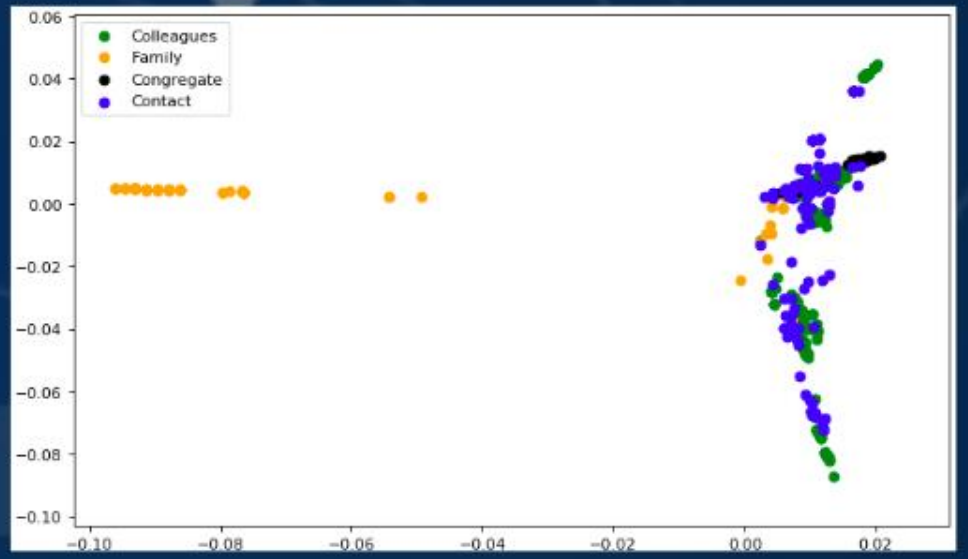
The impact on Spectral Clustering

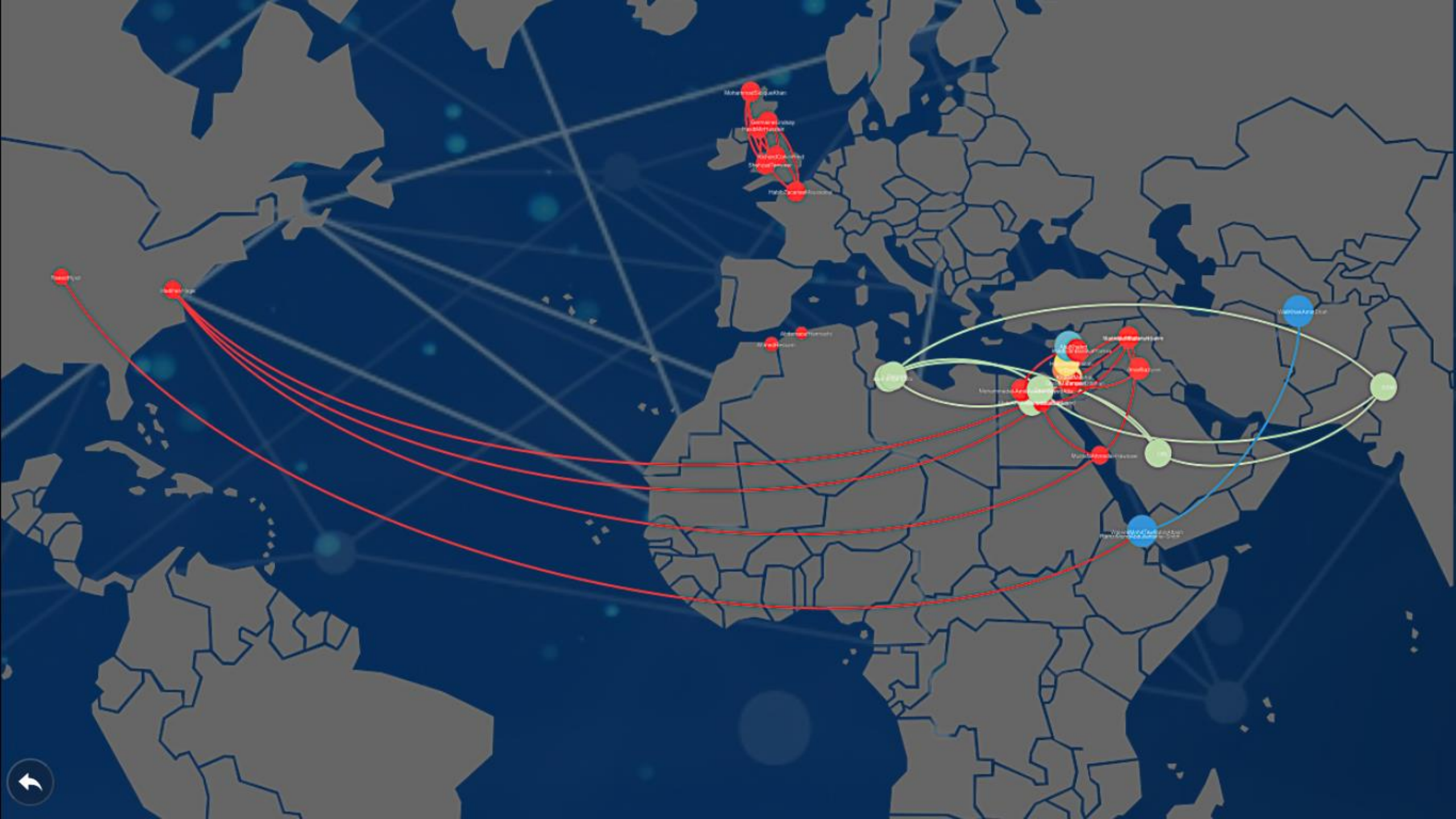
Re-performing the spectral clustering using k-means performed in the Milestones on the expanded network provided some interesting results:

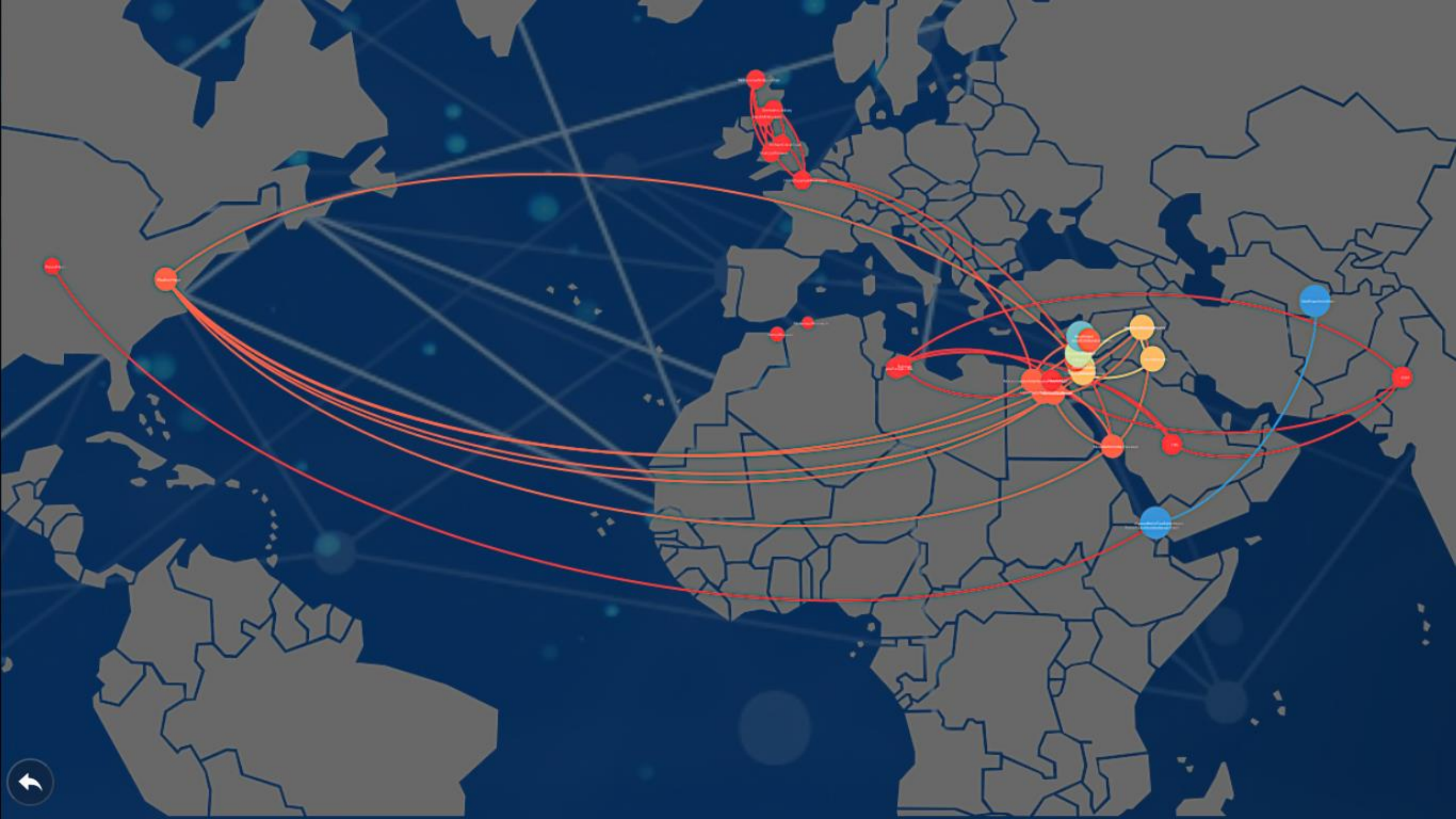
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Colleague	42%	16%	33%	9%
Family	35%	6%	0%	59%
Congregate	0%	94%	0%	6%
Contact	30%	25%	21%	24%

Cluster 1	Cluster 2	Cluster 3	Cluster 4
0%	33%	37%	30%
87%	9%	4%	0%
0%	100%	0%	0%
0%	57%	28%	16%









Analyzing signal diffusion across the network

- Performed a heat kernel on our Dual Graph and Expanded Dual Graph.
    - In order to see how exposing the true network changes the idea spread.
- Implemented with localized low-pass filtering of a graph with a dirac-delta signal on a particular node.
- Comparison of the diffusion rate is done visually.
    - Different parameters to select potentially the most dangerous terrorists.

Controlling Parameters

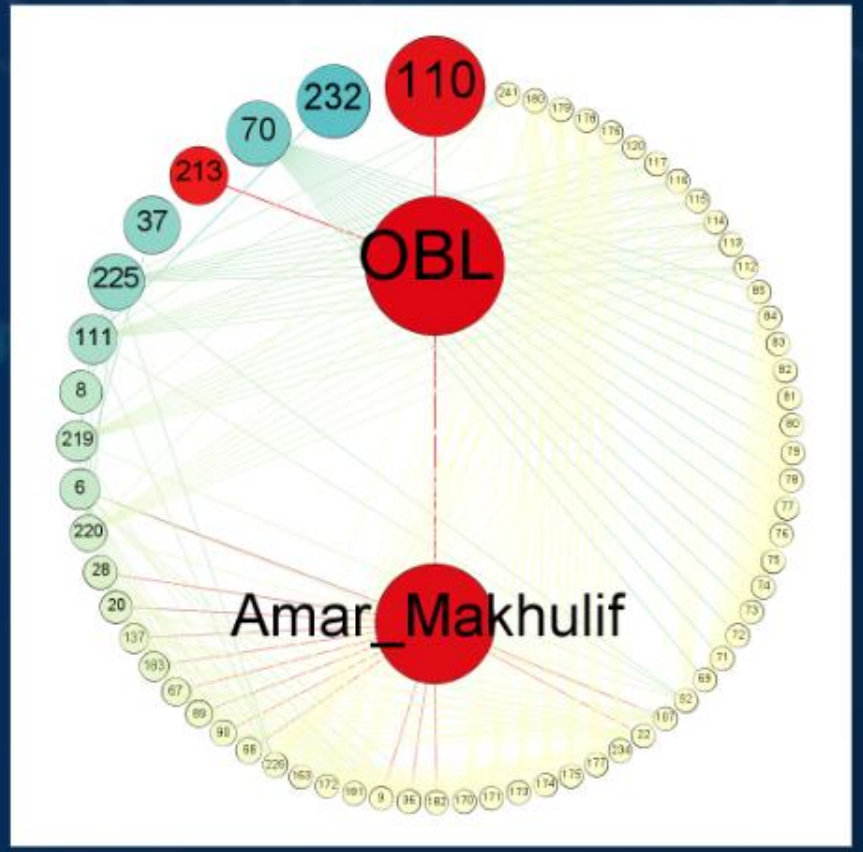
- **Degree:**
  - There are four highly connected terrorists.
  - Worked with 10 terrorists with highest degree.

- **Betweenness Centrality:**
  - Betweenness centrality of a ”node v” is the sum of the fraction of all-pairs shortest paths that pass through ”node v”.

$$c_B = \sum_{s,t \in V} \frac{\sigma_{s,t|v}}{\sigma_{s,t}}$$

where V is set of nodes,
$\sigma_{s,t}$ is the number of shortest (s; t)-paths,
$\sigma_{s,t|v}$ is the number of those paths passing through some node v other than s; t.

- The most ”wanted” terrorist since it can be interpreted as the center of the operation.



Controlling Parameters

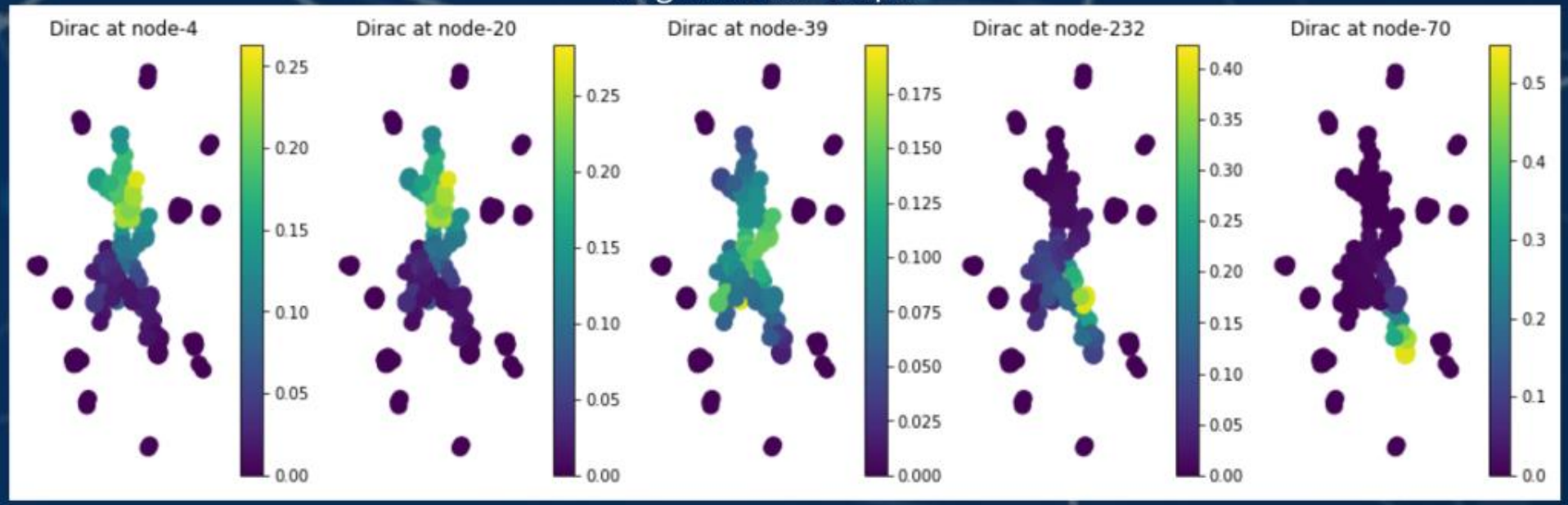
- **Articulation point:**
  - An *articulation point* is any node whose removal (along with all its incident edges) increases the number of connected components of a graph.
  - The leader of a branch of a terrorist organization.
    - Articulation points who create a second connected component with size smaller than 5 nodes are neglected.

Biggest Threats

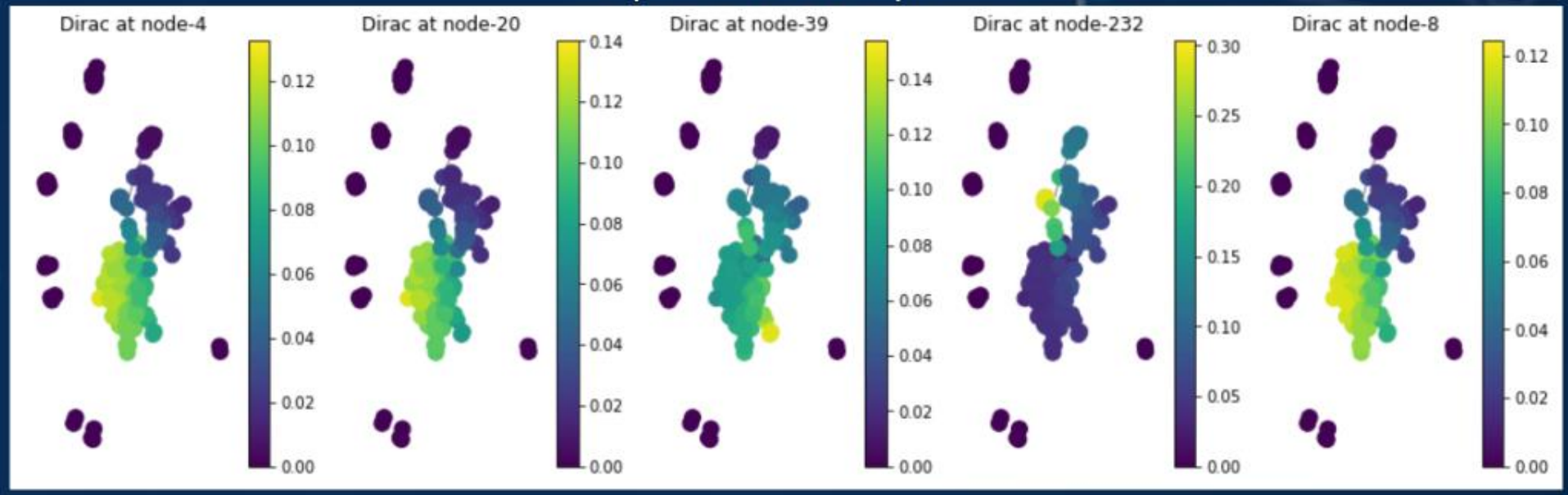
- Five terrorists for each graph that potentially constitute the biggest threat for the society.

Original Graph		Expanded Graph	
ID	Name	ID	Name
39	OBL	39	OBL
20	Mustafa Kamel	20	Mustafa Kamel
4	Amar Makhulif	4	Amar Makhulif
232	Zarqawi	232	Zarqawi
70	Abu Khaled	8	Fateh Kamel

Original Dual Graph



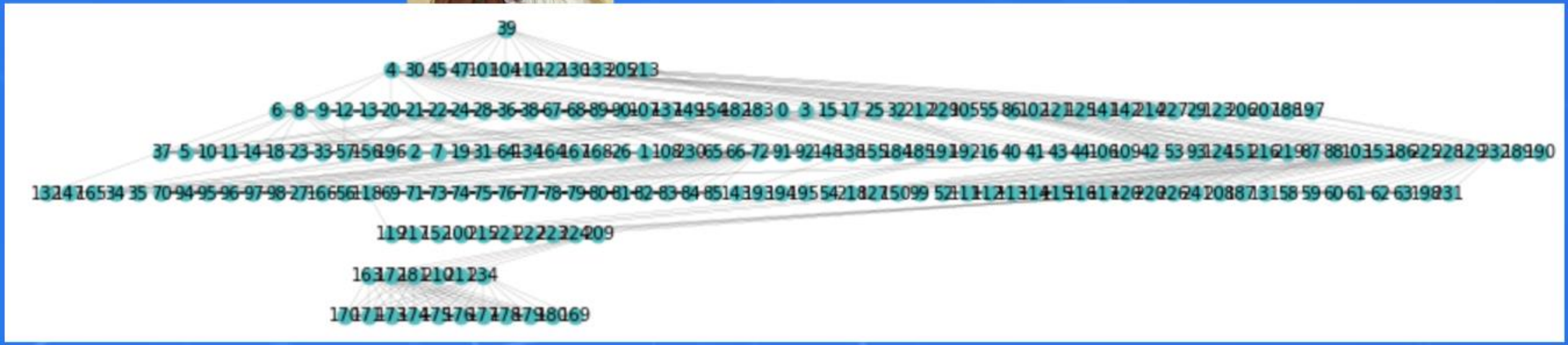
Expanded Dual Graph



Conclusion

- Predicted the existence of relations that are not publicly known
- Guessed the labels for new relationships by using transductive learning
  - Positive impact on the spectral clustering accuracy
- Consistency in threat diffusion between the original and the ‘true’ terrorist network
- Found the most dangerous terrorists in both networks by analyzing idea spread
  - The most central terrorist: Osame Bin Laden

Osame Bin Laden's connections in the expanded network



Q & A