

**ROXANNE**

**Real time network,  
text, and speaker  
analytics for  
combating  
organized crime**

# Case study

## Telephony data from a real case



This project has received funding from the European Union's Horizon 2020 Work Programme for research and innovation 2018-2020, under grant agreement n°833635

# Scenario

- Approximately 10 phone numbers are intercepted
- After clustering, the operator identifies the real identities of eight speakers of interest (one cluster for each)
  - We have some automatic procedures for this but this is not considered in this demo
- We want to better understand how they are connected
  - Talk to each other
  - Talk about each other, same places
  - Predict links
- Are there any other interesting people (unknown people linked with several criminals)?



# Case overview

- Wiretapped phone calls
- International drug case
- Total
  - 40k+ calls/SMS
  - ~2k tel. numbers
  - >1k speakers
- Relevant
  - ~ 200 calls
  - 90 tel. numbers
  - 70 speakers
- Calls are directed
- Audio is mono and undirected

start	end	tel no. source	tel no. target	name1	name2	Mentioned people	Mentioned Places	audio file
xx/xx/xx xx:xx:xx	xx/xx/xx xx:xx:xx	716	701	0	0			
xx/xx/xx xx:xx:xx	xx/xx/xx xx:xx:xx	701	716	0	0	13	A	
xx/xx/xx xx:xx:xx	xx/xx/xx xx:xx:xx	716	701	0	0			
xx/xx/xx xx:xx:xx	xx/xx/xx xx:xx:xx	1847	633	401	506	701,506		
xx/xx/xx xx:xx:xx	xx/xx/xx xx:xx:xx	2100	963	0	679			xxxxx.htm
xx/xx/xx xx:xx:xx	xx/xx/xx xx:xx:xx	1847	704	738	506		A,C	
xx/xx/xx xx:xx:xx	xx/xx/xx xx:xx:xx	755	1847	506	235	0		B
xx/xx/xx xx:xx:xx	xx/xx/xx xx:xx:xx	716	1847	506	838			



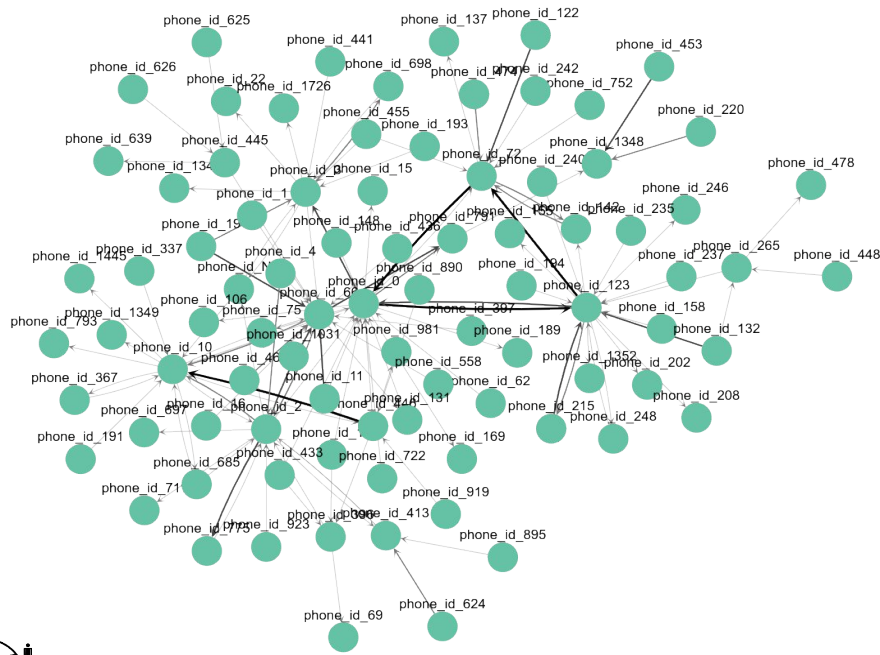
Multiple languages



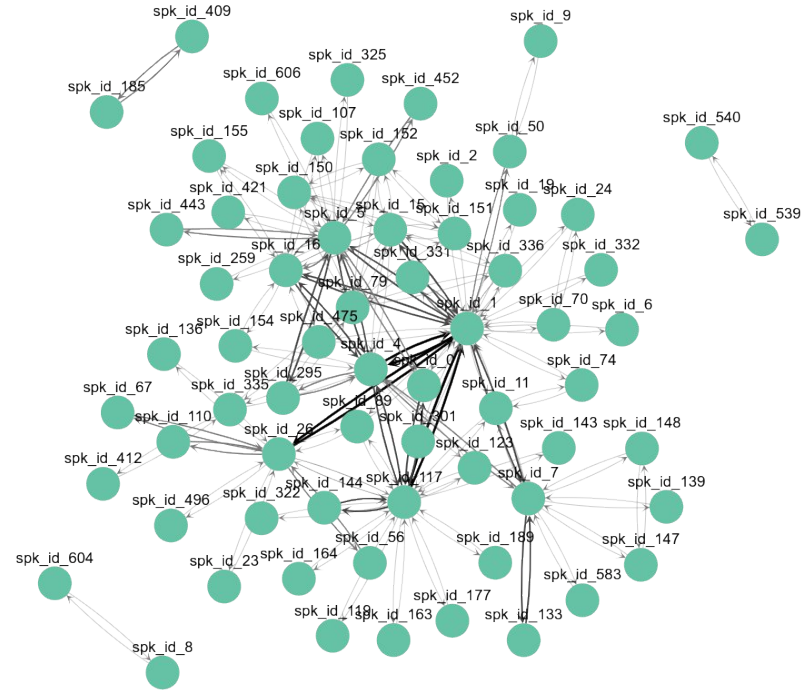
This project has received funding from the European Union's Horizon 2020 Work Programme for research and innovation 2018-2020, under grant agreement n°833635. © ROXANNE Consortium

# Visualization of the (criminal) network

- Phone number network

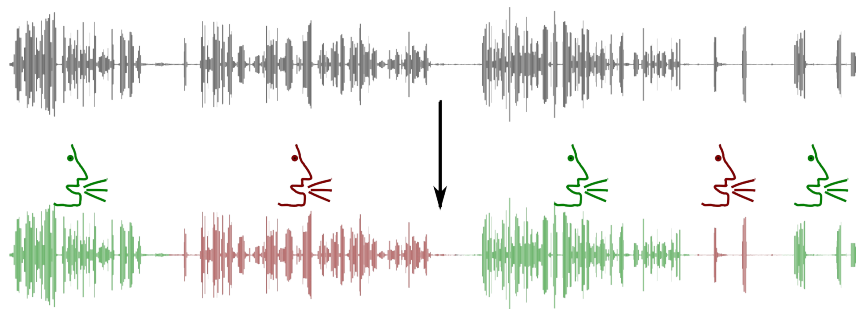


- Speaker network (ground truth)



# Diarization

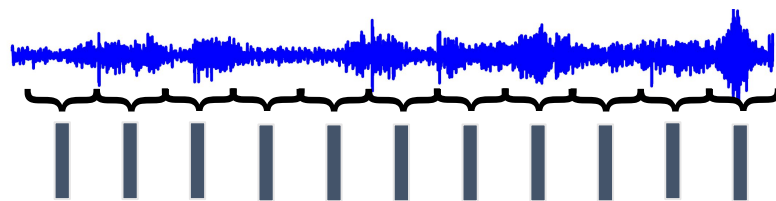
- The two sides of the telephone call is mixed into one mono channel
- Diarization segments the mono recording into speaker regions
- In our experiments, diarization is constrained to output two speakers



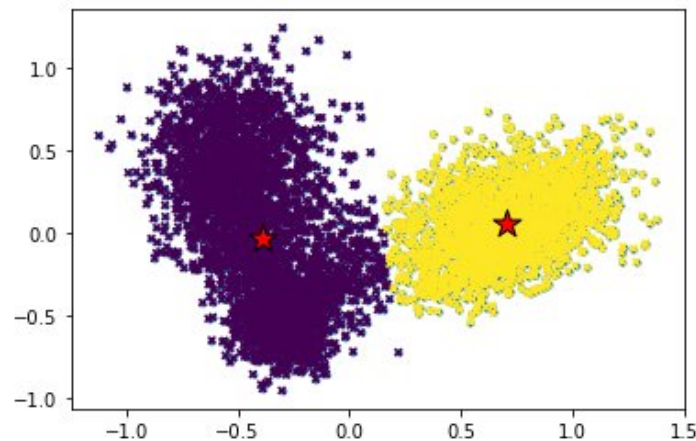
# Diarization procedure

1. Cut the recording into short segments (~0.75s)
2. Extract embeddings from the segments
3. *Cluster the embeddings*
4. Refine with more advanced model

Step 2



Step 3



# Anonymization through embeddings

- The word content in the recordings contains sensitive information that the LEA cannot share with other ROXANNE partners
- LEA provide speaker embeddings (*x-vectors*) to partners
  - Generally, word content cannot be inferred from the speaker embeddings

