UNIVERSITÀ DEGLI STUDI DI MILANO
DATA SCIENCE FOR ECONOMICS
Academic Year 2023-24

CRIMINAL NETWORK ANALYSIS:
Sicilian Mafia

PRESENTATION OUTLINE

- Research Question
- Dataset description
- Network description
- Random Network comparison
- Connectivity
- Transitivity
- Communities Detection
- Bosses analysis
- ML: Associates prediction
- Conclusion

"The strength of a family, like the strength of an army, lies in its loyalty to each other."
- Mario Puzo, The Godfather.

RESEARCH QUESTION

MAFIA CLANS: ORGANIZATION'S STRUCTURE AND BOSSES NODES

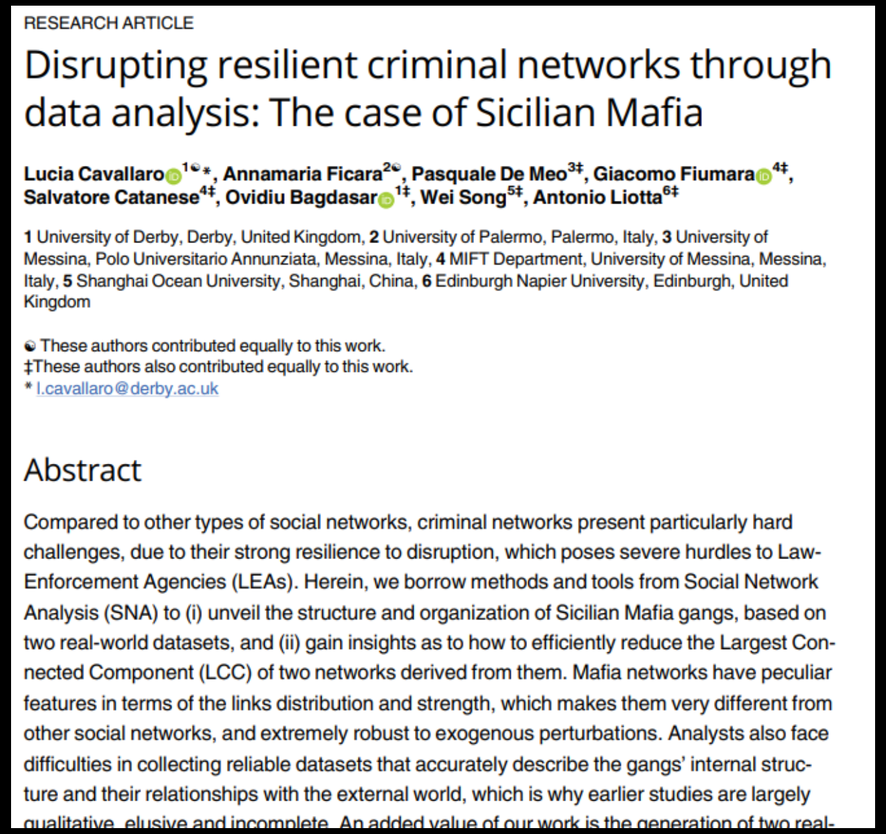
1. What is the structural organization of a Mafia clan, in regards to face-to-face meetings?
2. What is the role of bosses in the information flow?
3. To what extent can ML techniques help in predicting who are the clans' associates?

DATASET

"OPERAZIONE MONTAGNA": preliminary investigation order (Court of Messina, 2007)

Cavallaro, Lucia, et al. "*Disrupting resilient criminal networks through data analysis: The case of Sicilian Mafia*" (2020)

github.com/lcucav/criminal-nets



RESEARCH ARTICLE

Disrupting resilient criminal networks through data analysis: The case of Sicilian Mafia

Lucia Cavallaro[1☯*], **Annamaria Ficara**[2☯], **Pasquale De Meo**[3‡], **Giacomo Fiumara**[4‡], **Salvatore Catanese**[4‡], **Ovidiu Bagdasar**[1‡], **Wei Song**[5‡], **Antonio Liotta**[6‡]

1 University of Derby, Derby, United Kingdom, **2** University of Palermo, Palermo, Italy, **3** University of Messina, Polo Universitario Annunziata, Messina, Italy, **4** MIFT Department, University of Messina, Messina, Italy, **5** Shanghai Ocean University, Shanghai, China, **6** Edinburgh Napier University, Edinburgh, United Kingdom

☯ These authors contributed equally to this work.
‡These authors also contributed equally to this work.
* l.cavallaro@derby.ac.uk

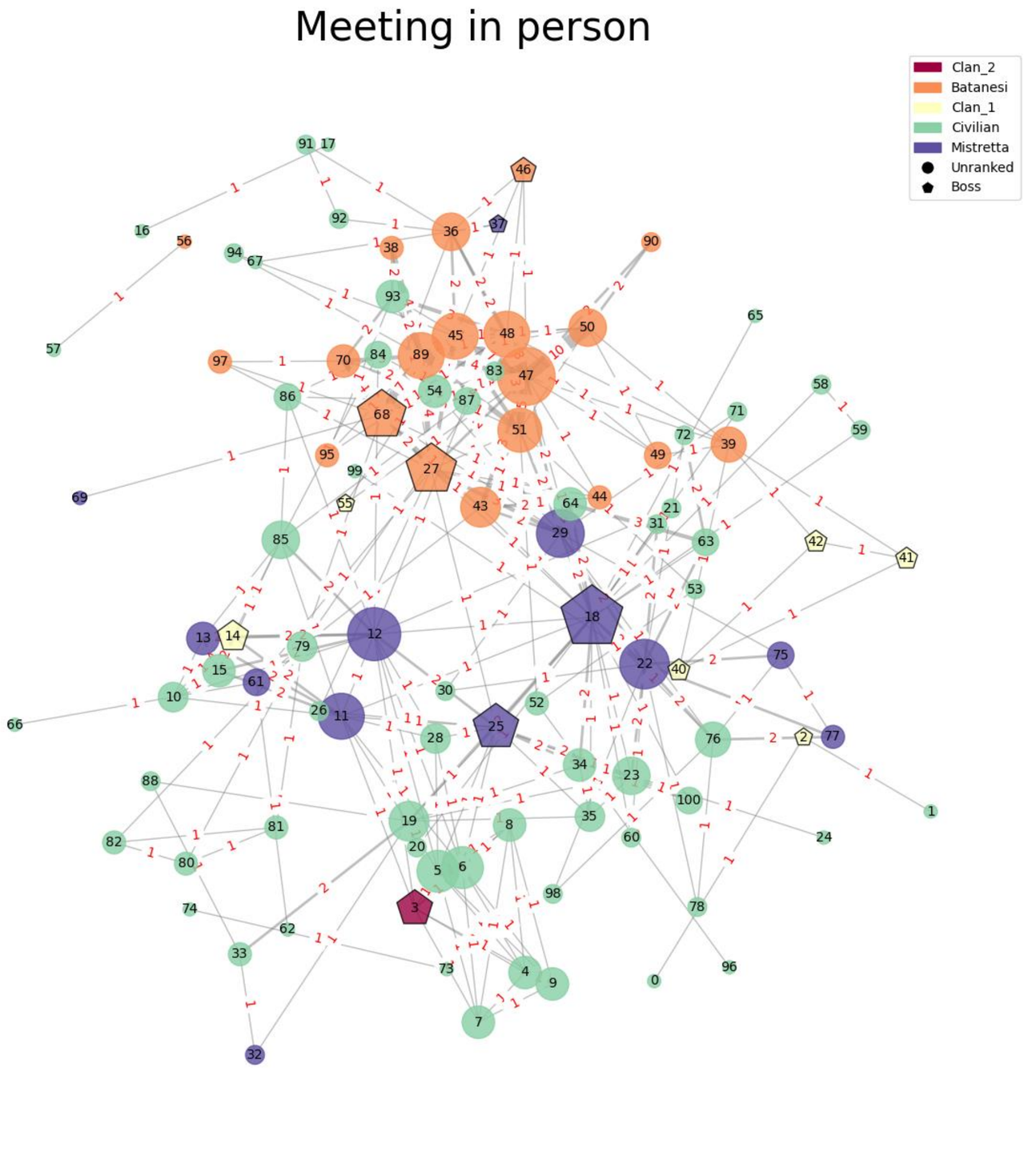
Abstract

Compared to other types of social networks, criminal networks present particularly hard challenges, due to their strong resilience to disruption, which poses severe hurdles to Law-Enforcement Agencies (LEAs). Herein, we borrow methods and tools from Social Network Analysis (SNA) to (i) unveil the structure and organization of Sicilian Mafia gangs, based on two real-world datasets, and (ii) gain insights as to how to efficiently reduce the Largest Connected Component (LCC) of two networks derived from them. Mafia networks have peculiar features in terms of the links distribution and strength, which makes them very different from other social networks, and extremely robust to exogenous perturbations. Analysts also face difficulties in collecting reliable datasets that accurately describe the gangs' internal structure and their relationships with the external world, which is why earlier studies are largely qualitative, elusive and incomplete. An added value of our work is the generation of two real-

NETWORK

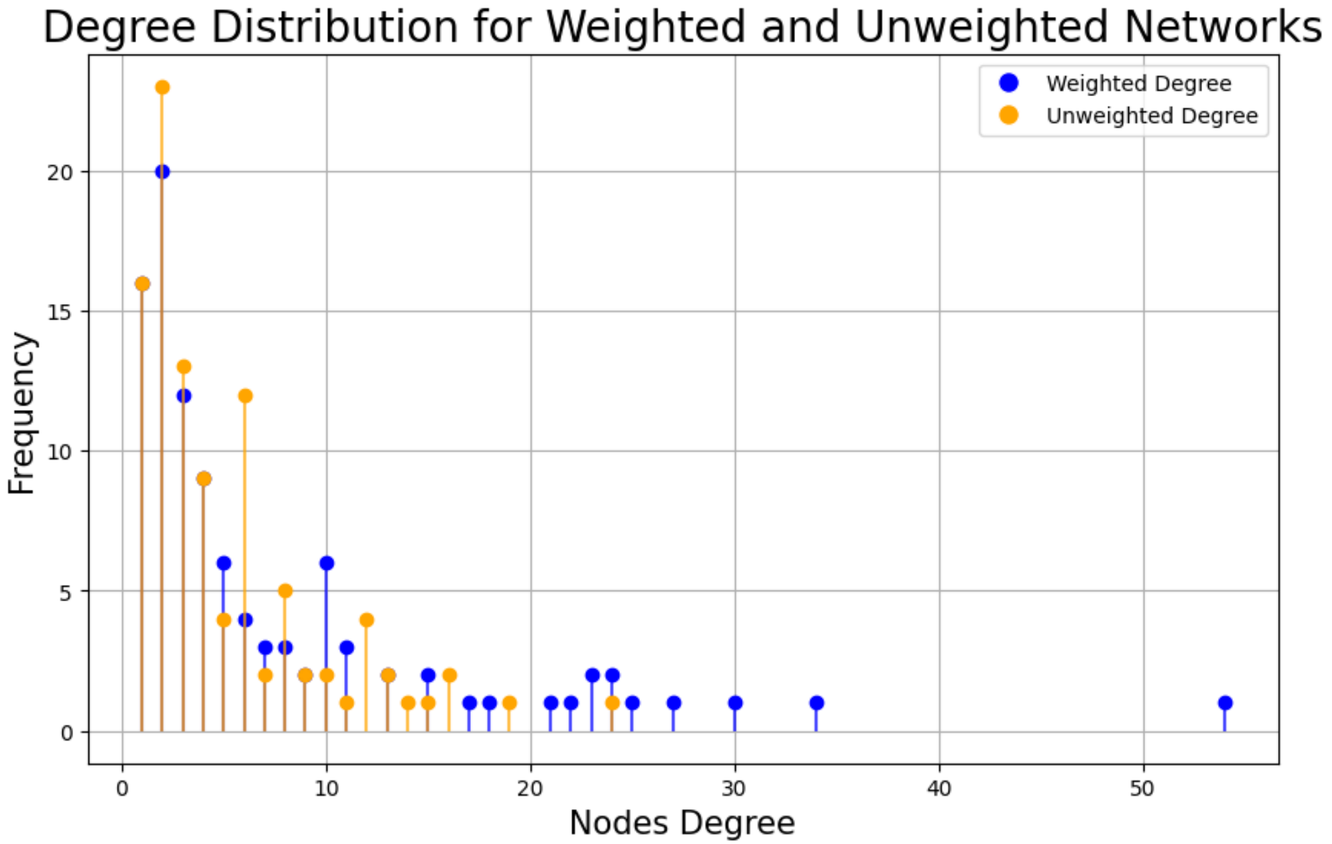
Meetings in person of surveiled suspects, with frequencies of meetings.

Nodes	**101**
Edges	**255**
Batanesi	20
Mistretta	13
Civilians	61
Other clans	7

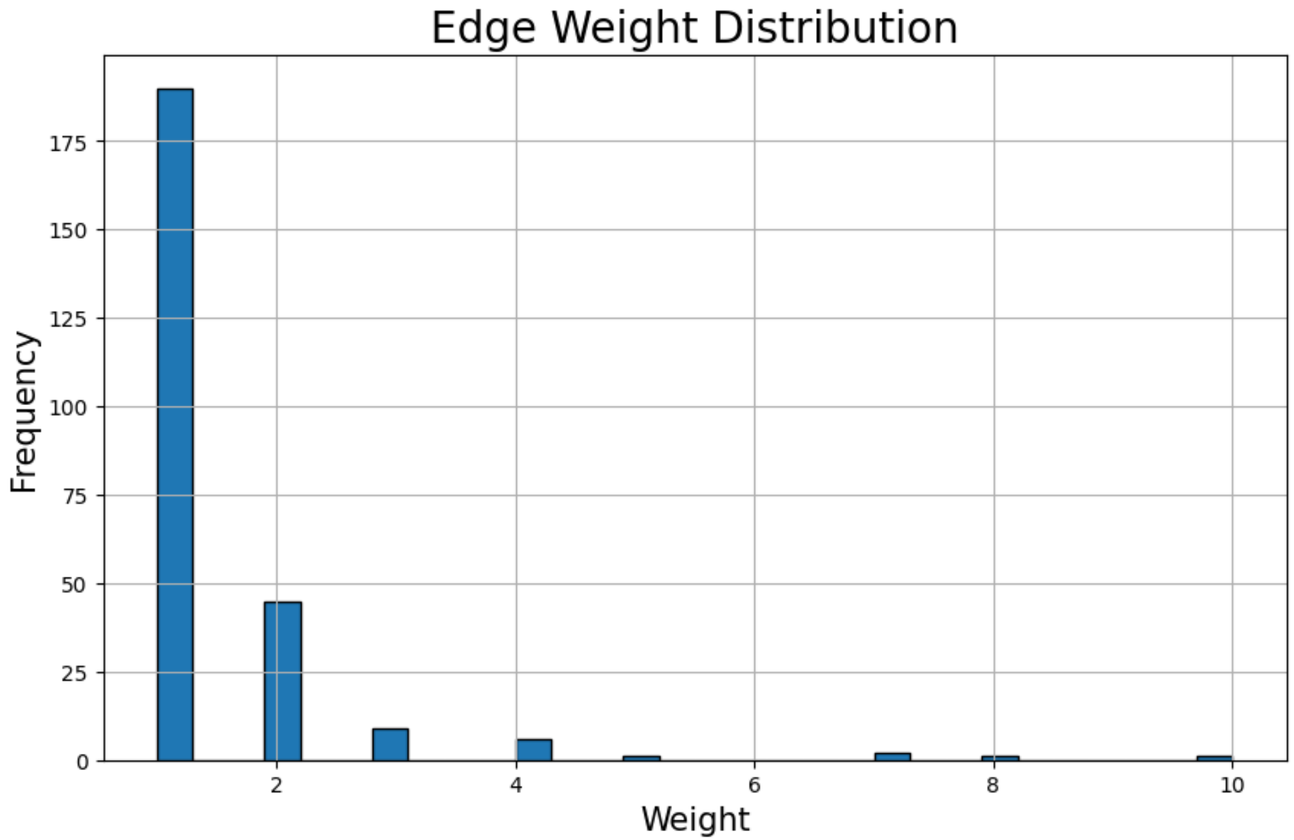


DEGREE DISTRIBUTION

METRIC	WEIGHTED	UNWEIGHTED
Density	**0.050**	**0.050**
Average Degree	**7.287**	**5.049**
Max Degree	**54**	**24**
Standad Dev	**8.677**	**4.462**



WEIGHTS DISTRIBUTION



Average Weight	**1.443**
Max Weight	**10**
Standard Deviation	**2.057**

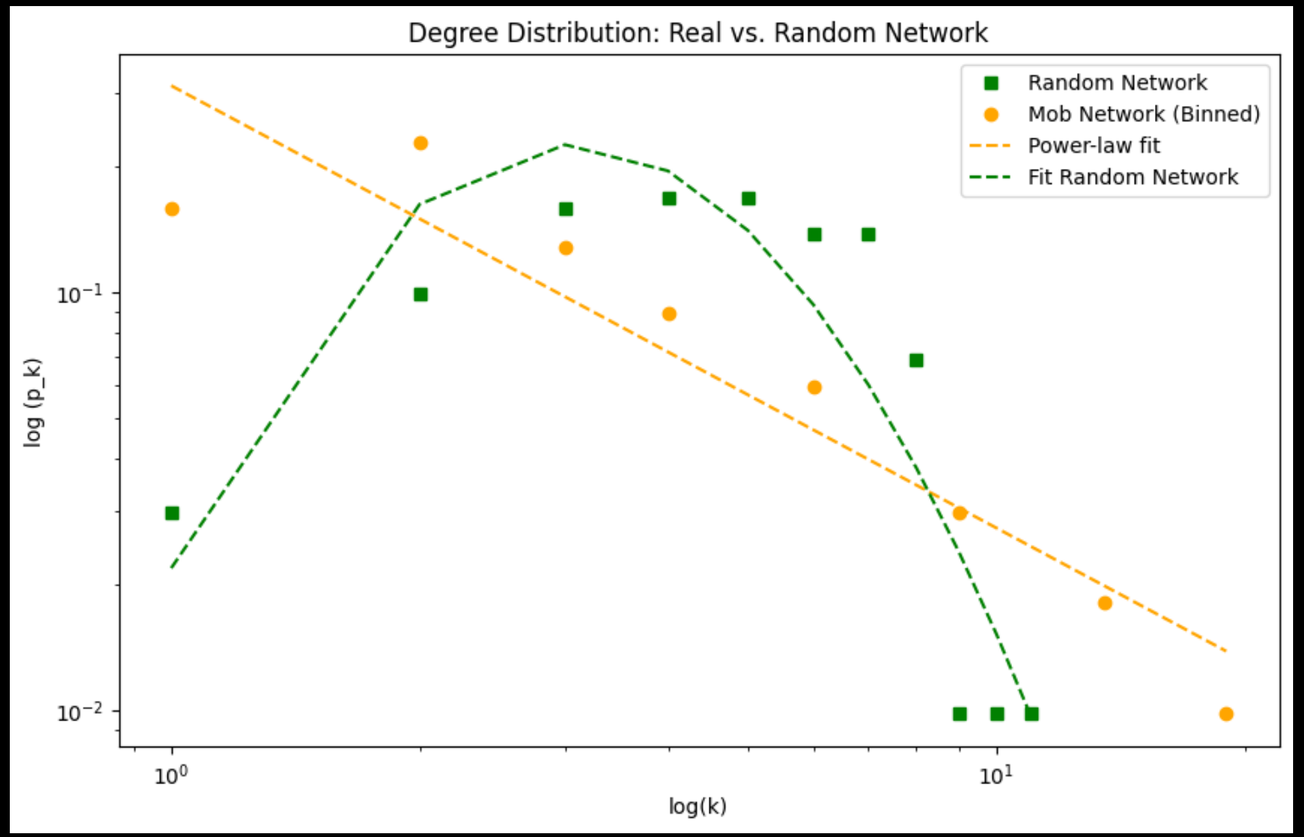
Majority of associates meet only once, presumably to minimize the exposure.

RANDOM NETWORK COMPARISON

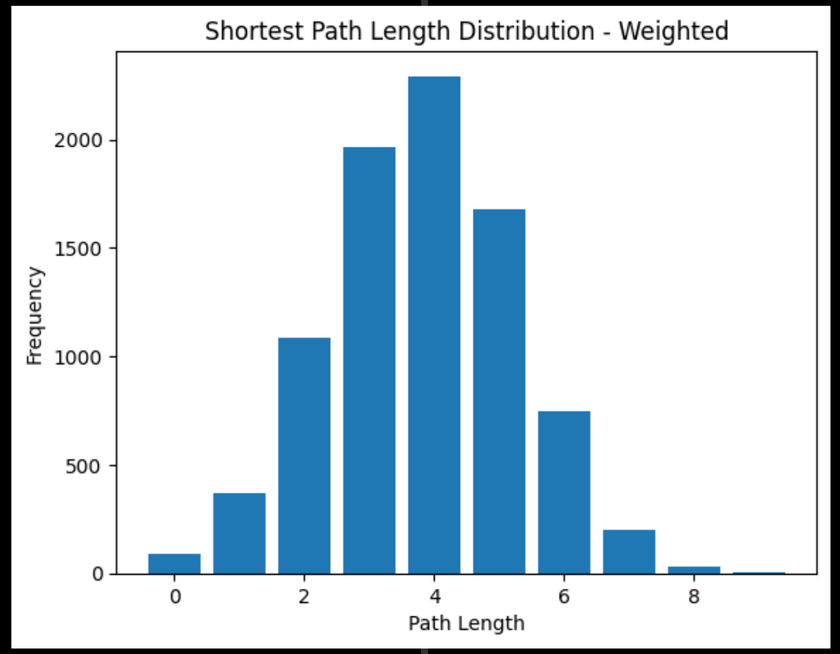
Erdos-Renyi model

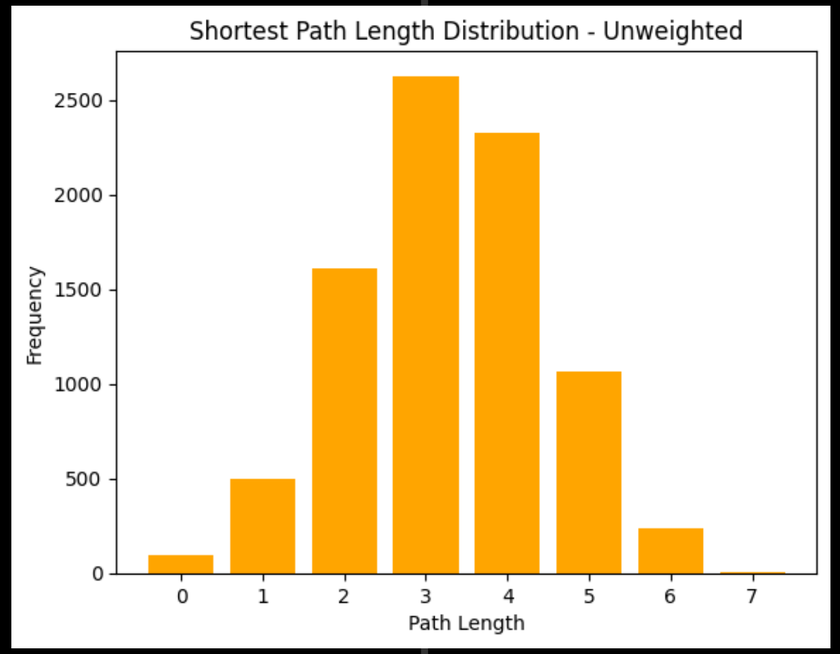
G(N = 101, p = 0.050)

The real network is skewed in the high degree area (right). There are more nodes with low degree and some with very high degree compared to the Random Network.



CONNECTIVITY





5 Connected Components

METRIC	WEIGHTED	UNWEIGHTED
LCC Order	**92**	**92**
LCC Size	**250**	**250**
Avg Shortest Path Length	**3.841**	**3.309**
Diameter	**9**	**7**
Number of Bridges	**9**	**9**

On average, it takes 4 people to deliver information across the organization, 7 people at most.

TRANSITIVITY

The real Network is more prone to clustering, common for Social Network, than the Random Retwork.

METRIC	WEIGHTED	UNWEIGHTED	RANDOM NETWORK
Avg Local Clustering Coefficient	**0.083**	**0.626**	**0.042**
Global Clustering Coefficient	**0.416**	**0.416**	**0.054**

DEGREE ASSORTATIVITY

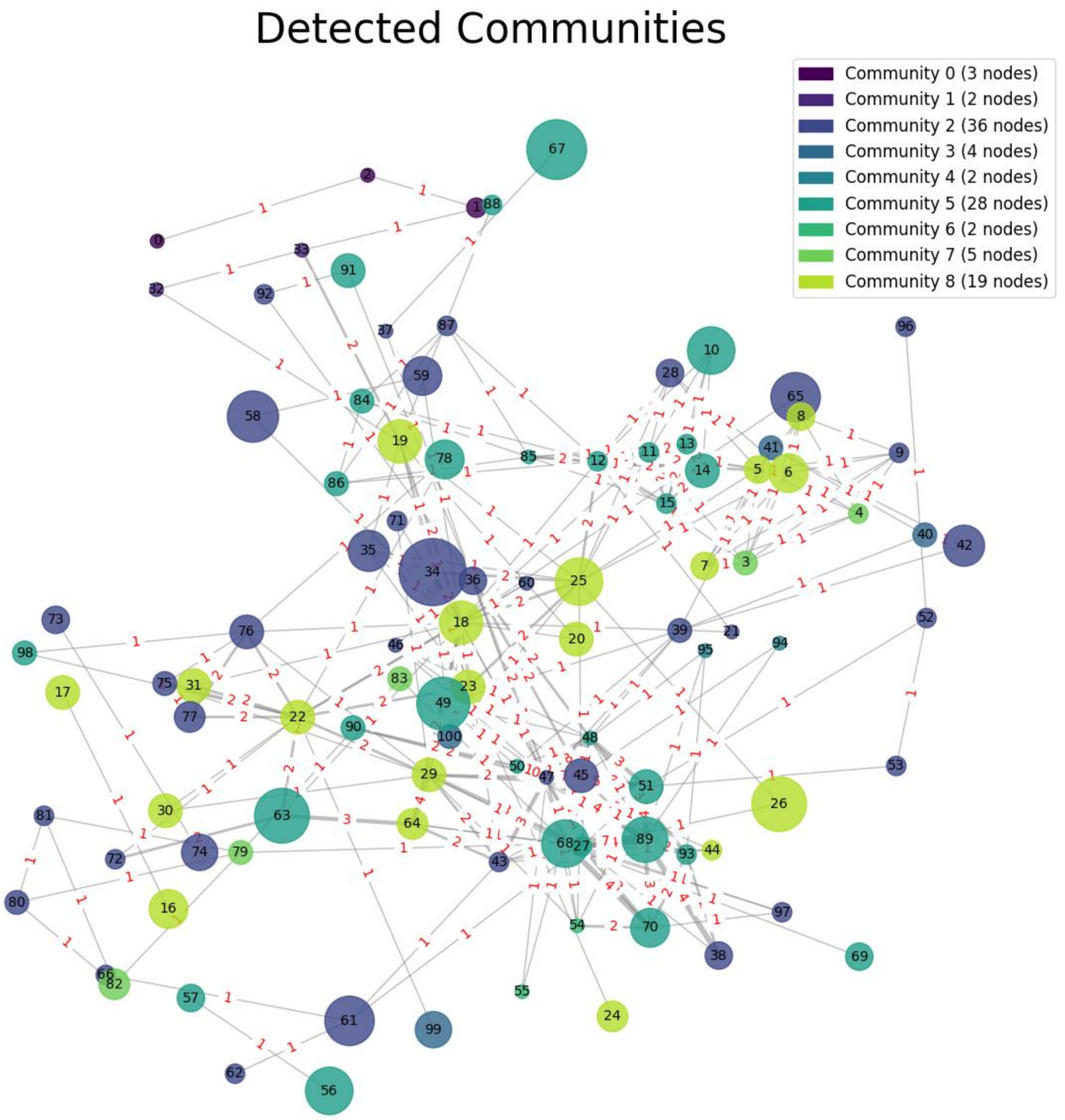
Weighted Network: **0.074**
Unweighted Network: **-0.043**

Low Assortativity means high connected nodes tend to link with low degree nodes.
Presumably, liutenants interacts more with underlyings than other liutenants.

COMMUNITIES

Community detection performed with the 2-Phases Louvain algorithm.

MODULARITY	WEIGHTED	UNWEIGHTED
Detected Communities	**0.572**	**0.589**
Clan	**0.246**	**0.194**



BOSS ANALYSIS

Four Bosses are fairly central in the two Clans.

Bosses 18 and 27 are the most centrals, taking more decisions in the organizations and presumebly yielding more power.

	Mistretta				Batanesi			
	18	**25**	**37**	***Avg***	**27**	**46**	**68**	***Avg***
Degree Centrality	30	17	2	*12.1*	27	4	25	*15.2*
Betweenness C.	0.418	0.082	0.018	*0.068*	0.132	0.018	0.201	*0.048*
Closeness C.	0.382	0.289	0.193	*0.257*	0.335	0.267	0.319	*0.256*
Eigenvector C.	0.064	0.019	0.005	*0.023*	0.231	0.054	0.219	*0.166*
Avg Shortest Path Length	2.357	3.120	4.652	*3.653*	2.685	3.370	2.826	*3.420*
Local Clustering Coeff.	0.020	0.044	0.0	*0.062*	0.05	0.145	0.048	*0.107*

Batanesi Clan display a higher average Local Clustering Coefficient than the whole Network (**0.083**).
The opposite is true for Mistretta Clan.

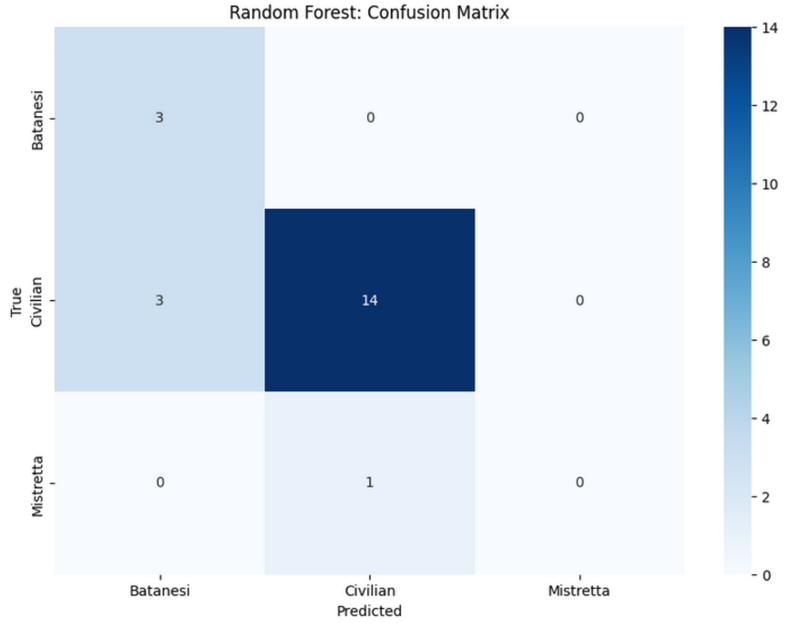
ML: ASSOCIATES PREDICTION

Encode the Graph with Node2Vec algorithm.

Apply Machine Learning algorithms to predict the clan label of the nodes.

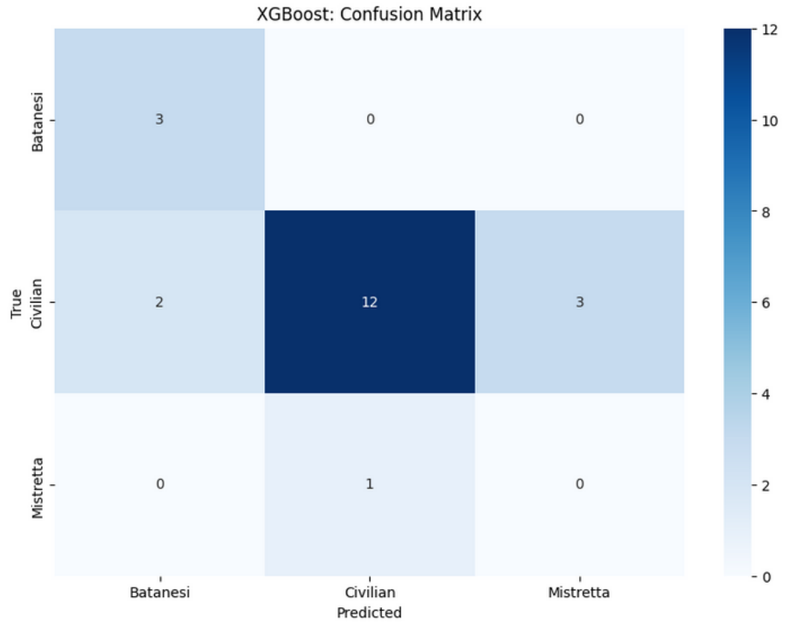
Random Forest Classifier

- Accuracy: 0.810



XGBoost

- Accuracy: 0.714



CONCLUSIONS

1. STRUCTURE

Low meetings frequency as a safeguard measure.
On average, 4 associates are enough to deliver orders and informations.
Network follows a power-law.

2. BOSSES

Bosses rely on intermediary nodes to send out orders, reducing risk of exposure.
For certain bosses centralities are inadequate metrics.

3. PREDICTION

ML techniques can help in predicting the clan associaties. The main issue is developing a dataset large enough (investigation challenges, “omertà”).