"Montagna Operation" Network Analysis: Unveiling the Sicilian Mafia Layout and Key Players

Telmo Ribeiro
Faculdade de Ciências da Universidade do Porto
Porto, Portugal
up201805124@fc.up.pt

Diogo Ferreira
Faculdade de Ciências da Universidade do Porto
Porto, Portugal
up201805258@fc.up.pt

ABSTRACT

The field of Network Science has witnessed remarkable advancements, benefiting various social sciences, including Criminology. By extracting network properties and conducting subsequent analyses, it becomes possible to comprehend the structure and dynamics of criminal organizations, thus shedding light on the core players involved.
This research article focuses on a specific anti-mafia operation called "Montagna", initiated by Italian law enforcement agencies with the objective of dismantling two families, namely "Mistretta" and "Batanesi". These families were suspected of colluding with the Sicilian Mafia, thereby implicating several influential entrepreneurs and high-ranking individuals in the politics of the city of Messina.
The datasets employed in this study were carefully curated by the authors, from sources with their original counterpart being, the judicial documents pertaining to the "Montagna" operation. Through meticulous preparation and integration of these datasets, we construct a comprehensive network that captures the relevant aspects of the criminal organization under investigation.

1. INTRODUCTION

Criminal networks present formidable challenges to law enforcement and society as a whole. Unraveling the intricacies of a specific criminal organization requires months or even years of meticulous operations, often placing the associated task force at significant risk. However, the rewards of such efforts are substantial, as they can lead to the cessation of illicit activities and the prevention of future criminal actions by the targeted group.
Traditional investigative techniques like stakeouts, wiretappings, and interrogations have long been employed to gather crucial data in these operations.
Nonetheless, in this article, we explore the possibilities of extracting valuable information using a purely "contact-based" approach.
Our investigation focuses on three distinct datasets: **meetings** and **phone calls**, which quantify the frequency of the respective contacts between pairs of individuals, and **contacts**, which flags alternative forms of relationships, such as individuals being observed together, engaging in transactions, or even taking a part in family ties.
By leveraging these datasets and guided by the known roles of the individuals, involved in the operation, our objective is to determine the potential insights that can be derived from this simplified network description method, when elevated through the use of the right metrics. We aim to extrapolate the results and engage in a discourse regarding the efficacy of incorporating network analysis or increasing its usage, as an auxiliary tool for task forces, thereby enhancing the likelihood of success in scenarios with much more information.

2. DATASETS DESCRIPTION

The research commenced with three distinct datasets, lacking a standardized format and exhibiting minimal cohesion among them [1].
Those datasets were utilized to extract our own, which were the basis for the research and were categorized as **meetings**, **phone calls**, and **contacts**, which then became edge lists. In addition to those edge lists, we were able to extract information regarding the known roles of certain individuals and their criminal status at the end of the trial, which became node lists. This information will contribute to the evaluation of the efficacy of our approach.

- The **meetings** dataset recorded physical encounters among suspected individuals, which were monitored through police stakeouts.
  Each edge in this dataset was weighted based on the frequency of meetings between the pair.

- The **phone calls** dataset documented intercepted phone conversations between individuals.
  Each edge in this dataset was weighted according to the number of times the pair engaged in phone calls.

- The **contacts** dataset focused on alternative forms of contact that could not be quantified using simple integers, such as family ties.
  Each edge in this dataset was weighted using a flag to indicate the presence of an alternative type of contact.

- The **role** property list provided information about the attributed roles of individuals within the organization. Some individuals were marked as "non-disclosured", and the reasons behind this classification may involve legal considerations, agreements, or the determination that the individual was not a member of the criminal organization associated with the two families.

- The **criminal status** property list documented the legal situations of certain individuals, including arrests.

Upon importing our data into **Gephi** [2], we constructed an **undirected network** that visualizes the known relationships between individuals.
It is important to note that since the **contacts** dataset measures contacts using a flag rather than precise edge weights, metrics relying on those may not be as accurate as desired. Nevertheless, one could argue that if this simplistic approach yields promising results, a more comprehensive dataset would likely achieve even greater success.

3. NETWORK MEASURES

3.1 Degree Distribution

As already mentioned, the network being researched is **undirected**, and as such, going forward we will not distinguish between in-degrees and out-degrees.

In this section, we will present and discuss its degree distribution.
In order to achieve such, we utilized one of the many **statistical tools** available on **Gephi**. This allowed us to determine the **average degree** within the network, which was found to be **4.542**, and additionally the degree distribution. Subsequently, we attempted to fit a power law model to it, as illustrated by the following image:

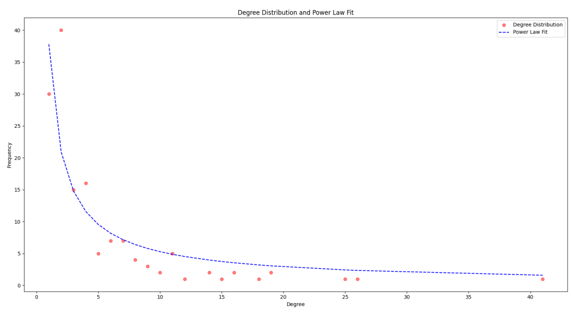

Figure 1: degree distribution - power law fitted (linear scale and non-normalized)

Our analysis yielded an estimated α of **0.855**.
In practical terms, such a **low α value** is not commonly present in social networks, where it usually ranges between **2** and **3**. That value would be an extreme case of the **ultra small world** phenomenon.

In the context of **power law distributions**, α relates to the scaling exponent and the degree of the **power law**. It governs the slope and describes the association, in this study, between the frequency of a node's degree and its occurrence. A **lower α value** indicates a slower decay and a higher concentration of nodes with higher degrees. Conversely, a **higher α value** corresponds to a faster decay and a more dispersed distribution of degrees. A **low value of α**, specifically less than **1**, signifies a divergent distribution characterized by numerous nodes with high degrees. This suggests the presence of some "powerful" hubs, something that we can find on the research network.

When we rank the network, meaning larger values imply larger nodes, by the node's degree we end up with the following visualization:

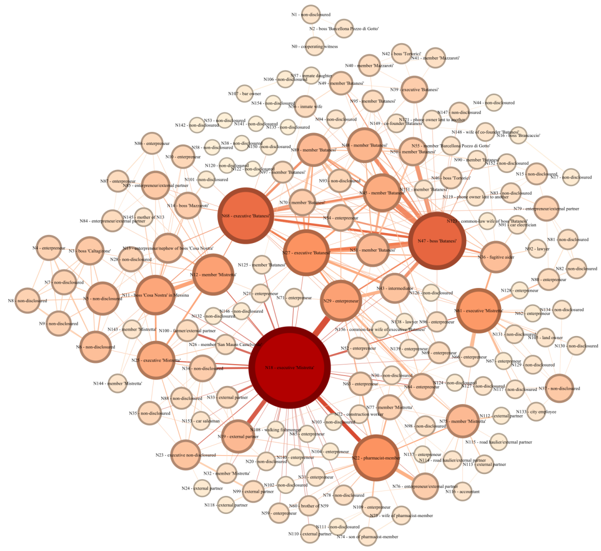

Figure 2: degree-based visualization - color and size

In order to ease the final discussion and give some statistical background, the following table showcases the rank-10 nodes respecting this metric:

Node	Degree
N18 - executive 'Mistretta'	41
N47 - boss 'Batanesi'	26
N68 - executive 'Batanesi'	25
N22 - pharmacist-member	19
N27 - executive 'Batanesi'	19
N61 - executive 'Mistretta'	18
N12 - member 'Mistretta'	16
N29 - entrepreneur	16
N11 - boss 'Cosa Nostra'	15
N25 - executive 'Mistretta'	14

Table 1: rank-10 - node degree

3.2 Clustering Coefficient

A node's clustering coefficient can be obtained through the formula [3]:

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

where e_i = #edges between the neighbors of i

Which then leads to the network's average being given by [3]:

$$C = \frac{1}{N}\sum_{i}^{n} C_i$$

Aimed again by **Gephi**'s statistical tools, we determined the network has a clustering value of **0.610**. In **social networks**, we would generally anticipate a relatively **high clustering coefficient**, when compared to a random model like **Erdős–Rényi's**.
However, the observed network exhibited an exceptionally high value, closely resembling what would be expected in a tightly connected group, such as movie actors [3]. This high value can be attributed to the specific nature of our study, which focuses on closely interconnected circles, namely a criminal investigation and the associated criminal organization.
A criminal investigation will quickly focus on known participants and their circles and almost as quickly dismiss guilt-free proven individuals.
A criminal organization, on the other hand, will try to be as undercover as possible, trying their best to not go public needlessly and, frequently, having the same individual fulfilling multiple roles (not hierarchical when regarding "Cosa Nostra").
The combination of both those factors may indeed explain, in part, the network's structure and some extreme results like this one.

3.3 Path Length and Diameter

The average path length can be calculated following the formula:

$$\bar{h} = \frac{1}{2E_{max}} \sum_{i,j \neq i} h_{ij}$$

where:
h_{ij} = distance between i and j
E_{max} = max number of edges

Gephi's calculations for this metric yield the value **3.136**, which is consistent with both the characteristics of this tightly interconnected circle and any social networks similar to it, as well as the random model **Erdős–Rényi**, which would anticipate an average path length of approximately $O(logn)$ where $n = \mathbf{155}$.
This tool also provides the network's **diameter**, yielding the value of **6** which is again a feature completely expected.

3.4 Connected Components

The **giant component** within the network encompasses a significant portion, precisely **89.68%**, of all nodes. Although this value somewhat aligns with the expectations provided by the characteristics of this network, considering its aforementioned attributes, one might expect a scenario where all nodes are part of the giant component.
However, it is important to note that certain **edges** may be **lost** due to **legal measures**, as mentioned previously. These measures could explain the lack of complete cohesion within the network, as well as strange behaviors like some higher-role individuals being not connected to the giant component.
While it is out of the scope of this article, it is worth mentioning that **link prediction algorithms** [4] could provide a more comprehensive understanding of the network. Such algorithms have the potential to alter this metric, along with other network characteristics, and offer a different perspective on the overall structure.

4. NODE CENTRALITY & LINK ANALYSIS

4.1 Betweenness Centrality

Betweenness centrality can be perceived as a measure of how much a node is involved in the connection of its peers.
To achieve this result, one could follow the given equation [3]:

$$C_B(i) = \sum_{j<k} \frac{g_{jk}(i)}{g_{jk}}$$

where:
g_{jk} = the number of shortest paths between j and k
$g_{jk}(i)$ = the number that i is on

As this role is closely related to the concept of brokerage, it is indeed used when **measuring the participation** of individuals in a criminal organization [5].

We can visualize the network when ranked by its betweenness centrality on the following image:

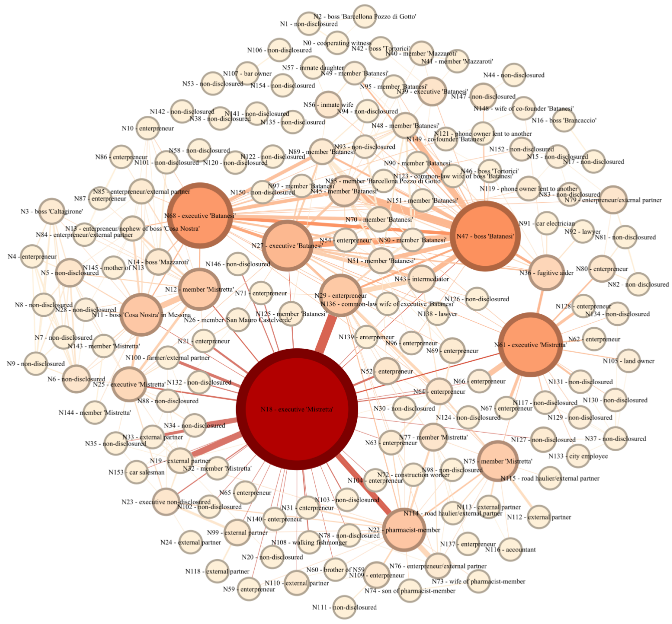

Figure 3: betweenness centrality-based visualization - color and size

The following table explores the rank-10 nodes ordered by betweenness centrality (normalized):

Node	Betweeness Centrality
N18 - executive 'Mistretta'	0.3537
N47 - boss 'Batanesi'	0.1690
N68 - executive 'Batanesi'	0.1520
N61 - executive 'Mistretta'	0.1469
N27 - executive 'Batanesi'	0.0999
N22 - pharmacist-member	0.0723
N11 - boss 'Cosa Nostra'	0.0699
N29 - entrepreneur	0.0693
N12 - member 'Mistretta'	0.0676
N75 - member 'Mistretta'	0.0631

Table 2: rank-10 - betweenness centrality

4.2 Closeness Centrality

Closeness centrality takes into account the average distance between an individual and the rest of the network.
It can be attained by the formula [3]:

$$C_C(i) = \frac{1}{\sum_j^N d_{ij}}$$

Closeness centrality, as betweenness centrality, is already used to **measure the importance** of an individual within an organization.

The next figure previews the network under its scope:

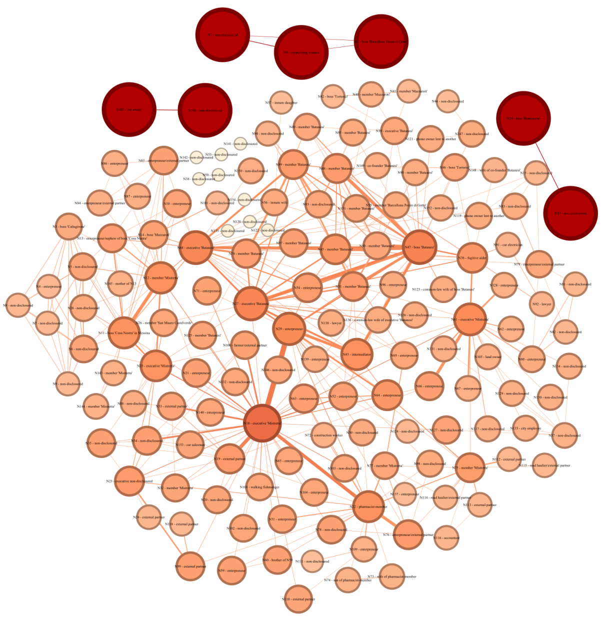

Figure 4: closeness centrality-based visualization - color and size

And again, we will explore the table with the rank-10 nodes ordered by closeness centrality (giant component only and normalized):

Node	Closeness Centrality
N18 - executive 'Mistretta'	0.5208
N27 - executive 'Batanesi'	0.4710
N47 - boss 'Batanesi'	0.4570
N29 - entrepreneur	0.4466
N68 - executive 'Batanesi'	0.4353
N43 - intermediator	0.4353
N61 - executive 'Mistretta'	0.4272
N12 - member 'Mistretta'	0.4157
N11 - boss 'Cosa Nostra'	0.4144
N64 - entrepreneur	0.4071

Table 3: rank-10 - closeness centrality

4.3 PageRank

PageRank is not a measure commonly seen applied in networks in order to extract the degree of involvement of a given individual in said network, although it is not a complete novelty also [6].
Nonetheless, if we picture the **PageRank value** as a measure of "importance" and assume that "important" people will circle themselves with similars, then we could make a case that this measure could shed light on insightful data.
That said, when employing PageRank, we need to deliberate on values for ϵ and p.
We assumed the standard value for p, that is, **0.85** would be a safe choice and as for ϵ, we deemed **0.001** as acceptable.

The next figure depicts the results found:

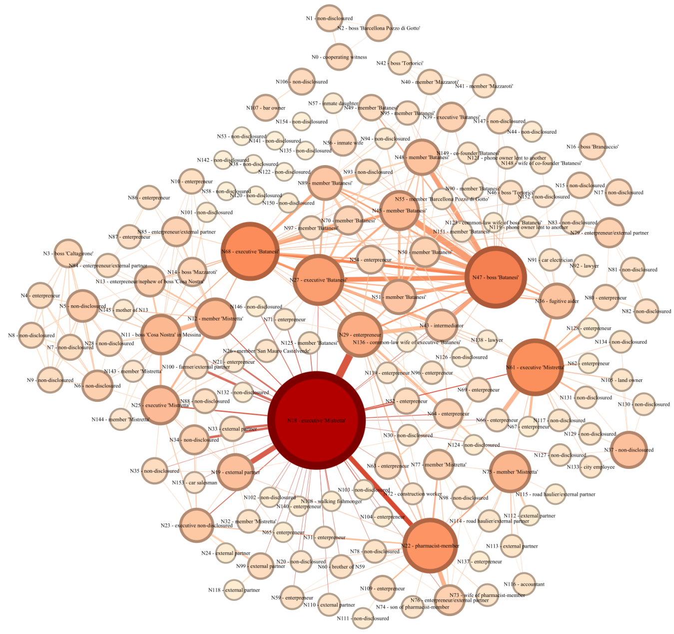

Figure 5: PageRank-based visualization - color and size

Finally, the rank-10 individuals by PageRank would be:

Node	PageRank
N18 - executive 'Mistretta'	0.0595
N47 - boss 'Batanesi'	0.0329
N68 - executive 'Batanesi'	0.0302
N61 - executive 'Mistretta'	0.0292
N22 - pharmacist-member	0.0277
N27 - executive 'Batanesi'	0.0249
N29 - entrepreneur	0.0205
N75 - member 'Mistretta'	0.0180
N12 - member 'Mistretta'	0.0177
N11 - boss 'Cosa Nostra'	0.0173

Table 4: rank-10 - PageRank

5. COMMUNITY STRUCTURE

Using **Gephi**, we employed the modularity measure to partition our dataset into distinct communities, taking into consideration only the **giant component**. **Modularity** is a property that quantifies the extent to which densely connected compartments within a system can be separated into separate communities or clusters that interact more within themselves than with other communities. In the end, we obtained a modularity value of **0.586**, using the standard range of [**0.3, 0.7**] as a good indicator of community structure, and the already expected outcome was verified.
The gripping results were the seven prominent communities found, **Gephi** allowed us to emphasize the connections between nodes, with each community represented by a unique color. Additionally, we observed variations in node size, reflecting their relative importance.
The chosen metric was **PageRank** since it was revealed as an exceptionally fitted measure.

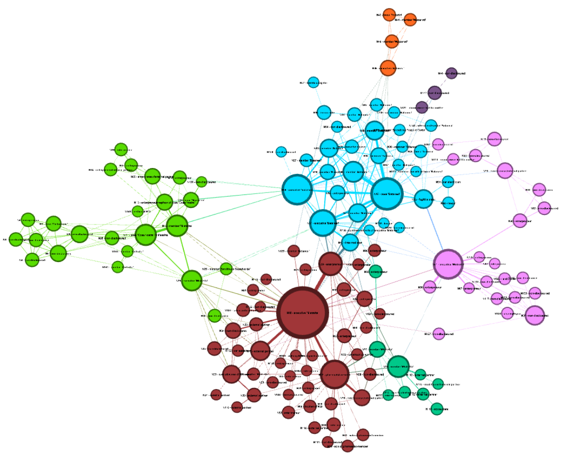

Figure 6: Communities overview

GREEN COMMUNITY: Analyzing the green community first, we can observe that the nodes with larger sizes are **"N11 - boss 'Cosa Nostra'"** and **"N12 - member 'Mistretta'"**, with a strong connection between them, evident from the thickness of the edge, suggesting frequent contact between them. Additionally, the **"N25 - executive 'Mistretta'"** stands out as a significant articulation with the red community. Overall, it is apparent that this community is primarily composed of members from the **'Mistretta' family** and a considerable number of entrepreneurs, suggesting the possibility of a branch carrying out a scheme.

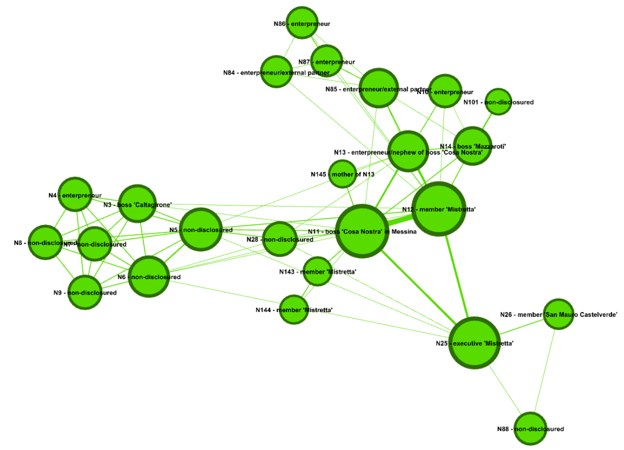

Figure 7: Green Community - PageRank

RED COMMUNITY: In turn, the red community also comprises members of the **'Mistretta' family**, with the node **"N18 - executive 'Mistretta'"** standing out significantly in size compared to other nodes within both the same community and the rest of the network.
We can see that this node has a strong connection with the nodes **"N29 - entrepreneur"**, another articulation between communities, and **"N22 - pharmacist - member"**. Similarly to the green community, we have the presence of multiple entrepreneurs and external partners, suggesting once again the existence of a scheme being carried out by this branch.
Furthermore, nodes with roles like **"N72 - construction worker"** display the amount of influence present in this family relating to schemes in public construction and real estate.

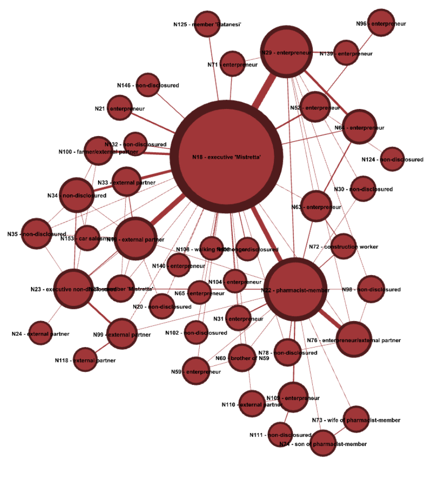

Figure 8: Red Community - PageRank

DARK-GREEN COMMUNITY: In the case of the dark green colored community, we have nodes such as **"N75 - member**

'Mistretta'" and **"N77 - member 'Mistretta'"** among others, taking the spotlight. The separation of this particular community from the others may be due to the fact of it having its own sphere of affiliated personnel, suggesting its own scheme, and the fact that some major nodes serve as articulations between the red and pink communities. In this particular community, nodes such as **"N114 - road hauler"** and **"N115 - road hauler"** may suggest once again some schemes related to public construction and real estate.

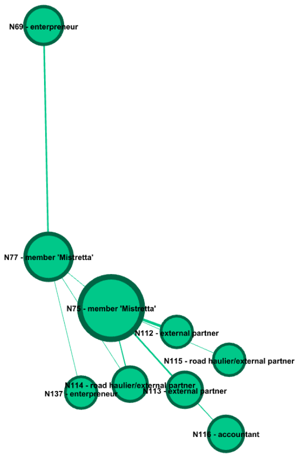

Figure 9: Dark-Green Community - PageRank

<u>*BLUE COMMUNITY:*</u> This community represents almost exclusively the members of the **'Batanesi' family**, with prominent nodes such as **"N47 - boss 'Batanesi'"**, **"N68 - executive 'Batanesi'"**, **"N27 - executive 'Batanesi'"**, and **"N43 - intermediator"**, which serves as an articulation to the previously mentioned red community.
It is worth noting the diversity of roles among the various nodes, reflecting the types of activities in which the **'Batanesi' family** is involved. Additionally, we can also observe the presence of the node **"N36 - fugitive aider"**, which articulates this community with the pink.

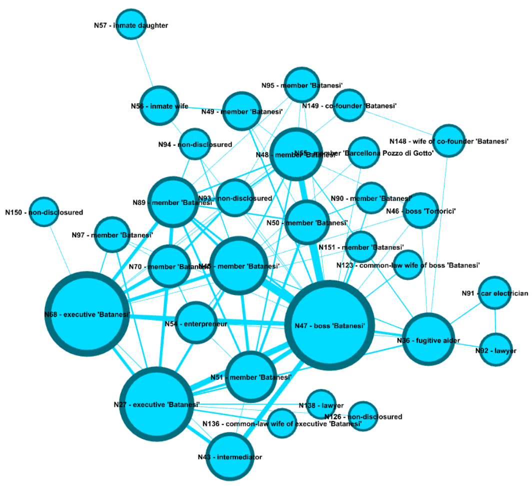

Figure 10: Blue Community - PageRank

<u>*PINK COMMUNITY:*</u> In this community, the spotlight goes to **"N61 - executive 'Mistretta'"**. Once again, we can observe that this community can be seen as an extension of the **'Mistretta' family**, further showcasing the sphere of influence of the former.
Again, the cloud of affiliated personnel suggests that this community is in control of its own schemes.

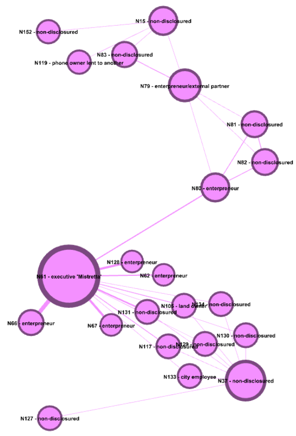

Figure 11: Pink Community - PageRank

<u>*ORANGE & PURPLE COMMUNITIES:*</u> These two smaller-sized communities reflect the extent of influence of the **'Batanesi' family**, particularly their connection with other less significant families (e.g., 'Mazaroti'), suggesting the potential existence of a partnership.

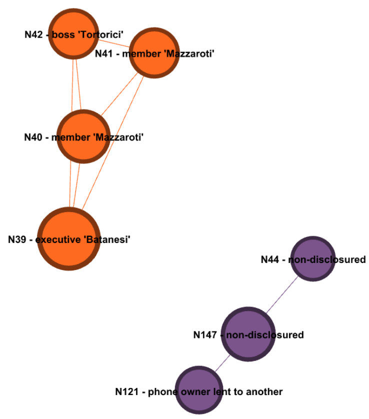

Figure 12: Orange & Purple Communities - PageRank

5.1 Overall community structure analysis

The obtained results for the **community structure** corroborate the real-life events, as they align with the findings of the **"Montagna Police Operation"**. This operation revealed that between 2003 and 2007, the **'Mistretta'** and **'Batanesi'** families had infiltrated various economic sectors, including public works, through a network of closely associated entrepreneurs linked to **"Cosa Nostra"**. These groups engaged in extortion and provided illicit protection to generate profits from public construction projects. Additionally, the investigation uncovered that the **'Mistretta'** family played a pivotal role as a mediator between **"Cosa Nostra"** families in Palermo and Catania, as well as other criminal organizations in Messina. Notably, both the **'Mistretta'** and **'Batanesi'** families maintained strong connections with other **"Cosa Nostra"** families, such as the **'Barcellona'** and the **'Caltagirone'** families, situated in the province of Messina. These charges were substantiated by multiple trials, resulting in lengthy prison sentences for the majority of individuals involved.[4]

6. CONCLUSIONS

As every section was used almost as a display of data, either by figures or tables, the conclusion will try to take the final discussion on the network's basic metrics, followed by its centralities and closing up on its communities.

Regarding the **degree distribution**, more precisely the extremely low α value obtained, although possibly explained through the reasons stated in that section, will still be treated as not expected and a considerable deviation. The use of **cumulative binning** [3] could be a technique that, more likely than not, would improve the result's quality. However, we would still argue the reasons presented, methods of improvement, and known characteristics of this network, are not enough to explain such a deviation.

Such is not the case when talking about the **clustering coefficient** and the **average path length/diameter** where is strongly believed the reasons given suffice to explain the values obtained in each of those sections.

About the topic of **connected components**, we emphasize the recommendation of the article on link prediction [4] which presents another structural overview of an extremely similar dataset

On **node centrality**, there are some arguments to be had. Node's **degree**, **betweenness centrality** and, less occasionally, **closeness centrality** are measures already applied in **Criminology** with high frequency and proven to yield results.
PageRank, as mentioned before, is being tested only much more recently.
Apart from some minor noise, all measures seem to agree on who the key players are, as such, we will consider that particular goal accomplished.
PageRank performed surprisingly well, knowing its only recent usage, in fact, we considered it the default network's overview from that point onward. We believe such behavior was successfully explained in its particular section, through the abstraction of "importance".
On the other hand, closeness centrality flattens the network, yielding less satisfactory results.
We justify such behavior on the network's domain. Being a short close connected circle, a node's distance to the "center" of the criminal web was almost constant, leading to the attained results.
We would speculate that the fact the majority of the key players are still individuals with great influence through the scope of closeness centrality, suggests the criminal investigation could have been mounted already focusing on those individuals and branching out from them.

On the subject of community structure, much has been argued about the real-life events and the communities themselves. Through **Louvain's algorithm** [7], **Gephi** was able to make reasonable distinctions on the communities. In fact, we can have a glimpse of the methods used by the families in their endeavors, the structural layout of the 'caporegime', and even the distinction between both subjects of study. The 'Mistretta' family seems to be less centralized, with prominent individuals, generally, executives ('capos'), controlling their own scheme and communicating with each other through articulations.
Nevertheless, the 'Batanesi' family seems much more centralized and compact, a strong community but with a smaller sphere of influence.
Finally, this simple "contact-based" approach was deemed able to extract plentiful information.
This mirror of the real-life events sure had shortcomings, but the majority of them regarded not being able to see the full picture, since for legal reasons, the full extent of data was not made available to the public. Due to the **Network Science emergence**, even the problem of working with incomplete datasets was softened. It would be reckless and naive to suggest a complete integration of such techniques in the task forces, without understanding the full capabilities and limitations of the former.
But as a tool, under the careful watch of an analyst, it is getting to a mature enough state that it can provide another perspective on the already tried and tested methods used by the leading investigators.

7. REFERENCES

[1] lcucav, "Network disruption." `https://github.com/lcucav/networkdisruption/`, 2020.

[2] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," 2009.

[3] "Network science - dcc/fcup." `https://www.dcc.fc.up.pt/~pribeiro/aulas/ns2223/`.

[4] F. Calderoni, S. Catanese, P. De Meo, A. Ficara, and G. Fiumara, "Robust link prediction in criminal networks: A case study of the sicilian mafia," *Expert Systems with Applications*, vol. 161, p. 113666, 2020.

[5] C. Morselli, "Assessing vulnerable and strategic positions in a criminal network," *Journal of Contemporary Criminal Justice*, vol. 26, no. 4, pp. 382–392, 2010.

[6] E. Budur, S. Lee, and V. S. Kong, "Structural analysis of criminal network and predicting hidden links using machine learning," 2015.

[7] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, p. P10008, oct 2008.

[8] L. Cavallaro, A. Ficara, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, and A. Liotta, "Disrupting resilient criminal networks through data analysis: The case of sicilian mafia," 03 2020.

[9] A. Ficara, L. Cavallaro, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, and A. Liotta, “Social network analysis of sicilian mafia interconnections,” in *Complex Networks and Their Applications VIII* (H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, and L. M. Rocha, eds.), (Cham), pp. 440–450, Springer International Publishing, 2020.

[10] A. Ficara, G. Fiumara, P. De Meo, and S. Catanese, “Multilayer network analysis: The identification of key actors in a sicilian mafia operation,” in *Future Access Enablers for Ubiquitous and Intelligent Infrastructures* (D. Perakovic and L. Knapcikova, eds.), (Cham), pp. 120–134, Springer International Publishing, 2021.

[11] A. Ficara, G. Fiumara, S. Catanese, P. De Meo, and X. Liu, “The whole is greater than the sum of the parts: A multilayer approach on criminal networks,” *Future Internet*, vol. 14, no. 5, 2022.

APPENDIX

All the images and complementary materials related to this project are available in the provided **.zip** file, named **"APPENDIX_NS-PROJECT.zip"**. This file contains additional resources that support the findings and analysis presented in this document. Please refer to the **.zip** file for access to the complete set of images and accompanying materials.