Network Features in Complex Applications

Lucia Cavallaro

College of Science and Engineering
University of Derby, Derby
United Kingdom

Supervisory Team

Director of Studies	Dr. Ovidiu Bagdasar
1st Supervisor	Prof. Fatih Kurugollu
External Supervisor	Prof. Antonio Liotta
	from *Free University of Bozen-Bolzano, Italy*

A submission in partial fulfilment of the requirements of the University of Derby for the award of the degree of Doctor of Philosophy

September 2021

To my aunt Laura

Declaration

This is to declare that the work stated in this dissertation was done by the author, and no part of the dissertation has been submitted in a dissertation form to any other university or similar institution. No human or animal participation have been included in this research and the research presented in this dissertation has been ethically approved. The author confirms that this work has not been accepted in substance for any degree and is not concurrently presented for any degree other than Doctor of Philosophy (PhD) being studied at the University of Derby and that appropriate credit has been given within the thesis where reference has been made to the work of others. Parts of this dissertation have previously appeared in the papers listed in the primary list of publications.

Lucia Cavallaro
University of Derby
2021

Abstract

The aim of this thesis is to show the potential of Graph Theory and Network Science applied in real-case scenarios. Indeed, there is a gap in the state-of-art in combining mathematical theory with more practical applications such as helping the Law Enforcement Agencies (LEAs) to conduct their investigations, or in Deep Learning techniques which enable Artificial Neural Networks (ANNs) to work more efficiently. In particular, three main case studies on which evaluate the goodness of Social Network Analysis (SNA) tools were considered: (i) Criminal Networks Analysis, (ii) Networks Resilience, and (iii) ANN topology.

We have addressed two typical problems in dealing with criminal networks: (i) how to efficiently slow down the information spreading within the criminal organisation by prompt and targeted investigative operations from LEAs and (ii) what is the impact of missing data during LEAs investigation.

In the first case, we identified the appropriate centrality metric to effectively identify the criminals to be arrested, showing how, by neutralising only 5% of the top-ranking affiliates, the network connectivity dropped by 70%.

In the second case, we simulated the missing data problem by pruning some criminal networks by removing nodes or links and compared these networks against the originals considering four metrics to compute graph similarities. We discovered that a negligible error (*i.e.,* 30% difference from the real network) was detected when, for example, some wiretaps are missing. On the other hand, it is crucial to investigate the suspects in a timely fashion, since any exclusion of suspects from an investigation may lead to significant errors (*i.e.*, 80% difference).

Next, we defined a new approach for simulating network resilience by a probabilistic failure model. Indeed, while the classical approach for removing nodes was always successful, such an assumption was not realistic. Thus, we defined some models simulating the scenario in which nodes oppose resistance against removal.

Once identified the centrality metric that on average, generates the biggest damage in the connectivity of the networks under scrutiny, we have compared our outcomes against the classical node removal approach, by ranking the nodes according to the same centrality metric, which confirmed our intuition.

Lastly, we adopted SNA techniques to analyse ANNs. In particular, we moved a step forward from earlier works because not only did our experiments confirm the efficiency arising from training sparse ANNs, but they also managed to further exploit sparsity through a better tuned algorithm, featuring increased speed at a negligible accuracy loss. We focused on the role of the parameter used to fine-tune the training phase of Sparse ANNs. Our intuition has been that this step can be avoided as the accuracy loss is negligible and, as a consequence, the execution time is significantly reduced. Yet, it is evident that Network Science algorithms, by keeping sparsity in ANNs, are a promising direction for accelerating their training processes. All these studies pave the way for a range of unexplored possibilities for an effective use of Network Science at the service of society.

Contents

Acknowledgements

At first, I would like to acknowledge the valuable guidance of my Director of Studies Dr. Ovidiu Bagdasar and of Prof. Antonio Liotta, who welcomed me at the beginning of this experience. They supported me and taught me what really matters other than giving me all the professional support I needed.

I would also like to thank Prof. Fatih Kurugollu that is my first supervisor, but also Prof. Ashiq Anjum for being part of my supervisory team. Their support was essential to pursue my PhD journey.

I would also like to express sincere thanks to Prof. Pasquale De Meo, who believed in me since my Bachelor Degree, and Prof. Giacomo Fiumara for their abundant patience and expertise.

I would like to acknowledge Dr. Decebal Mocanu for sharing his expertise in Sparse Artificial Neural Networks and providing data and code.

I am also very thankful to Ms. Tamsin Espinosa, for her fundamental help and feedback in improving my Academic English skills.

To Melissa, which acted like another big sister, and to Laura, Maryleen, Enrico, and Francis for their precious friendship and support in a foreign country.

Finally, my express gratitude goes to my beloved partner Tony, all my family and close friends, especially to Riccardo and Mary for supporting me in difficult times.

I am immensely grateful to everyone that has supported me along this journey.

List of Publications

Journal Papers

- **L. Cavallaro**, S. Costantini, P. De Meo, A. Liotta, G. Stilo, Network connectivity under a probabilistic node failure model, IEEE Transactions on Network Science and Engineering (under review). Pre-print: https://arxiv.org/abs/2006.13551
- A. Ficara, **L. Cavallaro**, F. Curreri, G. Fiumara, P. De Meo, O. Bagdasar, W. Song, A. Liotta, Criminal networks analysis in missing data scenarios through graph distances, *PLoS ONE*, 16(**8**) (2021) 1-18. https://doi.org/10.1371/journal.pone.0255067
- **L. Cavallaro**, A. Ficara, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, W. Song, A. Liotta, Disrupting resilient criminal networks through data analysis: The case of Sicilian Mafia, *PLoS ONE*, 15(**8**) (2020) 1-22. https://doi.org/10.1371/journal.pone.0236476
- **L. Cavallaro**, O. Bagdasar, P. De Meo, G. Fiumara, A. Liotta, Artificial neural networks training acceleration through network science strategies, *Soft Computing* 24 (2020) 17787–17795. https://doi.org/10.1007/s00500-020-05302-y

Conference Papers

- **L. Cavallaro**, M. Grassia, G. Fiumara, G. Mangioni, P. De Meo, V. Carchiolo, O. Bagdasar, A. Liotta, Relations between Entropy and Accuracy trends in Complex Artificial Neural Networks, In: Complex Networks and Their Applications X (under review).
- **L. Cavallaro**, A. Ficara, F. Curreri, G. Fiumara, P. De Meo, O. Bag-

dasar, A. Liotta, Graph comparison and artificial models for simulating real criminal networks, In: R. Benito, C. Cherifi, H. Cherifi, E. Moro, L. M. Rocha, M. Sales-Pardo (Eds.), Complex Networks and Their Applications IX, Springer International Publishing, 2021, pp. 286-297. https://doi.org/10.1007/978-3-030-65351-4_23

- A. Ficara, **L. Cavallaro**, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, A. Liotta, Social Network Analysis of Sicilian Mafia Interconnections, In: H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, L. M. Rocha (Eds.), Complex Networks and Their Applications VIII, Springer International Publishing, 2020, pp. 440-450 https://doi.org/10.1007/978-3-030-36683-4_36
- **L. Cavallaro**, O. Bagdasar, P. De Meo, G. Fiumara, A. Liotta, Artificial Neural Networks Training acceleration through Network science strategies, In: Y. D. Sergeyev D. E. Kvasov (Eds.), Numerical computations: Theory and algorithms, Springer International Publishing, 2020, pp. 330-336. https://doi.org/10.1007/978-3-030-40616-5_27

Book and Chapters

- **L. Cavallaro**, O. Bagdasar, P. De Meo, G. Fiumara, A. Liotta, Graph and network theory for the analysis of criminal networks, In: G. Fortino, A. Liotta, R. Gravina, A. Longheu (Eds.), Data Science and Internet of Things: Research and Applications at the Intersection of DS and IoT, Springer International Publishing, 2021, pp. 139-156. https://doi.org/10.1007/978-3-030-67197-6_8

Dataset Published

- **L. Cavallaro**, A. Ficara, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, W. Song, A. Liotta, Criminal Network: The Sicilian Mafia. "Montagna Operation", Zenodo 0.0.1 (July 2020) [Data set]. http://doi.org/10.5281/zenodo.3938818

Reactions in Press

TG3 Leonardo https://tinyurl.com/TG3Leonardo (Coming Soon, Oct 2021)

TGR Piemonte https://tinyurl.com/TG3Piemonte

TGR Bolzano https://tinyurl.com/TG3Bolzano

UoD News https://tinyurl.com/UoDNews

UniBZ News https://tinyurl.com/UniBZNews (Translated in: ENG, ITA, and DE)

Eurekalert.org https://tinyurl.com/EurekAlertOrg

Galileo net.it https://tinyurl.com/GalileoNewsWebsite

Ansa.it https://tinyurl.com/AnsaPhysMathIta

ZME Science.com https://tinyurl.com/ZmeScienceCom

List of Abbreviations

AI Artificial Intelligence

ANNs Artificial Neural Networks

BC Best Connected

HPC High Performance Computing

LCC Largest Connected Component

LEAs Law Enforcement Agencies

ML Machine Learning

MLP Multi-Layer Perceptron

OSNs Online Social Networks

SNA Social Network Analysis

SET Sparse Evolutionary Training

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Context

In this section, the research context on which this thesis was born is described.

For years Graph Theory has been used for mathematical purposes only. Recently, however, there is a growing interest in its application to real-world problems and this is how Network Science was born. Indeed, it is a branch of Graph Theory that also borrows techniques from other scientific fields (*e.g.*, Maths, Physics, Statistics, Computer Science, Sociology, etc.). More specifically, there is a shortage of studies on some application domains in which Network Science could be nevertheless a powerful tool in adding relevant information on the way in which some problems are addressed.

This is where this thesis was born. In particular, the focus is mainly on two topics: Criminal Networks and Artificial Neural Networks (ANNs). Those fields may appear to be unrelated to each other, but the goal of the research herein presented is to pave the way to combine them together. Indeed, this work addresses four of the most significant problems in those areas.

The first two of these problems, herein addressed with Social Network Analysis (SNA) tools, are directly related to Criminal Networks. In fact, the aim was to (i) identify the most critical points able to cause the greater damage in the information spreading process within a criminal organisation (Chapt. 3), and to (ii) evaluate the damage that Law Enforcement Agencies (LEAs) has in having a partial knowledge of the criminal organisation (Chapt. 4).

The remaining two are preliminary studies that have among their possible future extensions a strong connection with the application domain of criminal networks. The first one is a new model that was born to simulate more faithfully the real world behaviours in terms of network resilience features (Chapt. 5). The latter is an investigation on Artificial Neural Network topologies aimed at their optimisation in terms of complexity and costs (Chapt. 6).

1.2 Problem Description and Research Questions

This section outlines the problem description jointly with a brief preview of the research questions and related outcomes.

The first requirement was to obtain reliable criminal networks on which conduct some preliminary analyses to have a clearer scenario of how those networks are composed and what their topology is.

Consequently, taking the information from juridical acts, two real criminal graphs were created. On the graphs, the shortest path length (*i.e.*, the minimum number of acquaintances for a person needed to reach another individual in the network) and the degree distribution (*i.e.*, the probability distribution of the number of directly connected acquaintances individuals have over the whole network) were analysed to understand how the connections (*i.e.*, links or edges) were made. Next, with knowledge it was possible ranking the suspects (considering them as the nodes, or vertices, of the graphs) according to different centrality metrics (*i.e.*, ranking strategies of individuals importance within the network) with the aim to identify the most suitable for the specific scenario under scrutiny; namely: (i) *Degree centrality*, (ii) *Betweenness centrality*, (iii) *Katz centrality*, and (iv) *Collective Influence*. Generally speaking, Degree centrality helps identifying network hubs (*i.e.*, a node with a number of incident links that greatly exceeds the average). Nonetheless, contrary to other types of networks, criminal networks communications are not necessarily mediated via hubs that are more visible and, thus, more vulnerable [1, 2]. Hence, from the analyses emerged that the Betweenness Centrality performs well on the analysed criminal networks: by neutralising (*e.g.*, arresting) only 5% of the top-ranking affiliates, the network connectivity (*i.e.*, Largest Connected Component, or LCC, size) dropped by 70%.

Of course, another relevant problem in conducting investigations on criminal organisations relates the missing data from LEAs side. This is why this thesis was developed also on this direction; indeed, a partial knowledge of the connections

among suspects may seriously compromise the investigation success. Hence, to simulate the missing data scenario, several criminal networks were pruned and the obtained networks were compared with the original ones. In this way, it was possible to highlight the differences between original networks and their pruned counterparts by an evaluation of their similarities (resp., dissimilarities) [3]. The evaluation has been computed by the use of (i) *Adjacency*, (ii) *Laplacian*, (iii) *normalised Laplacian*, (iv) *Spectral*, and (v) the *Root Euclidean* distances.

This analysis was twofold; from one side, it was analysed how much the lack of information in terms of unavailability of proves during the investigations on specific criminals (*i.e.*, nodes) affect the knowledge of the overall network. From the other side, it was evaluated the lack of investigation details represented by a fragmentary knowledge among suspects' connections (*i.e.*, links). This last point is caused by several issues that may arise during the investigative process. To mention a few: (i) incompleteness, due to the covert nature of criminal organisations; (ii) incorrectness, caused by either unintentional data collection errors or intentional deception by criminals; (iii) inconsistency, when the same information is collected into law enforcement databases multiple times (or in different formats) and, thus, the connections have to be discarded from the analysis; (iv) time constraints (*i.e.*, there is a limited time available for eavesdropping specific devices or people), and so on.

As expected, an unsatisfactory knowledge of the suspects habits cause higher damages to the investigation comprehensiveness, because, if an individual cannot be under surveillance, then all their acquaintances information are lost as well. However, what was surprising from the analyses herein conducted was that even a large portion of unavailable connections (*i.e.*, 10%) still make satisfactory accurate knowledge of the overall networks under scrutiny possible (*i.e.*, there is only a 30% non-similarity from the original network compared with the 80% of dissimilarity in the case of missing nodes).

Having reached a clearer understanding of how to deal with criminal networks, it became a crucial point to discover why and how those specific kind of networks are so resilient. Thus, another problem arose. In Network Science, indeed, the concept of resilience is a well-known issue. However, there is a shortage of studies about methodologies able to translate this aspect into mathematical models. Indeed, generally speaking, all the simulations for nodes removal are defined as always successful; in a nutshell, the node is just removed from the network. Sometimes, this approach does not reflect what happens in the reality.

For this reason, investing some time in defining a new approach was needed: the probabilistic failure model (Chapt. 5), which associates each node with its probability to survive from a failure. When the node survival probability is zero (herein addressed as "Benchmark" analysis), then the model coincides with other models already introduced in the literature [4, 5]. In other words, the model takes into account the possibility of a node to fail (from the inside) or to be attacked (from the outside) and, thus, when the threshold is exceed the node is consequently removed from the graph.

In Network Science, this behaviour can be translated in a huge amount of possible real scenarios. As an example, nodes may represent power grids that, if affected of malfunctions, turn off and can cause blackouts in specific areas. In this case, peripheral power-grids are generally more fragile and, thus, their failure is more likely. On the other side, power-grids in the city centre are typically more resilient to failures and, thus, their strength is expected to be higher, by construction.

Unsurprisingly, in those kind of well-engineered networks (equipped with means to resist to hacking and to hardware or software failures [4]) the resilience level is stronger and well-documented. Thus, those networks are herein used as a litmus test when the model is applied on different network topologies. Indeed, another possible application domain of this model is in social networks, in which the resilience can be seen as a resistance from some people to be convinced (or get influenced) by some ideas (*e.g.*, marketing suggestions, political fashion, etc.) or peculiar abilities of suspects in hiding their trails and making LEAs collection of proof to arrest them difficult. However, due to lack of data availability this specific context was discarded in this preliminary study with the aim of consolidating the model before applying it on criminal networks.

The model was considered in two variants: (i) *Uniform* nodes survival-to-failure probability; and (ii) *Best Connected* (BC) in which survival probability proportional to node degree. The computation was performed by using five of the most popular centrality metrics (*Degree*, *H-Index*, *Coreness*, *Eigenvector*, and *Katz*), and evaluated by the *effectiveness* (*i.e.*, the drop in the spectral radius λ_1 after node removal) and *coverage* (*i.e.*, the reduction c of the LCC size of a graph), on eight real–world graphs. The spectral radius, in fact, governs a broad range of spreading processes in the graphs, such as the diffusion of an infection [6, 7, 8, 9], malware propagation [10, 11], or the dissemination of fake news in Online Social Networks (OSNs) [12, 13, 14], whereas the LCC size is widely used to quantify the resilience of a natural or artificial system described by a graph [4].

What emerged from the analysis was that the node degree can generally be used to cause the biggest drop in both λ_1 and c, especially in graphs deriving from human interactions/collaborations. Compared with conventional methods, the probabilistic model exhibits significant differences (ranging from 0% to 83%), highlighting the benefits of this method.

As previously asserted, the aim of this thesis is to show some of the potential of Network Science by detecting the most relevant features to be applied in different application domains, in particular in criminal networks. A growing field is the use of Artificial Intelligence (AI) for managing most of the aspects of our lives: home automation, Big Data analysis, prediction and classification problem resolutions, and so on. Unfortunately, mainly due to the lack of data availability, it is almost impossible nowadays to apply such techniques on criminal networks as they are.

Nevertheless, it is still important to understand how Artificial Neural Network (*i.e.*, one of the famous AI tool) techniques work and have deep knowledge of those strategies. Generally speaking, an ANN is used as a black-box in which some inputs are provided and some outputs are returned as results. However, it is not known how those networks work and why they are so accurate (when properly modelled). In addition, those networks perform better with resource-consuming configurations, which require High Performance Computing (HPC) power, and a huge amount of input information is needed. Thus, it became necessary to streamline those structures without (or, at least, with a negligible) accuracy loss.

For those reasons, the last aspect that was covered in this thesis is a preliminary study on how ANNs can be optimised to speed up the execution time and reducing the computational cost. Starting from Sparse ANNs, theorised with the Sparse Evolutionary Training (SET) framework by *Mocanu et al.* [15], an analysis on this model was herein conducted, more specifically on a parameter that adjusts the connections among neurons during the training process. For the sake of brevity, the novelty introduced in SET was to have a sparse ANN instead of the classic fully connected one. The reduction of connections makes the network lighter and allows simulations to be run even on laptops. Training a neural networks involve an update of the weights. During this process, in SET, a fraction of the less significant weighted links is rewired. In this way, the network has the same number of links during the overall process but, at the same time, the connections are refreshed.

In this thesis, then, the selection criterion of this rewiring fraction size was analysed on three publicly available datasets. It emerged that this step can be avoided

(*i.e.*, no rewiring phase is needed in most of the cases) still keeping the accuracy loss negligible. This discovery speeds up the execution time and the computational cost is reduced consequently. Then, the promising outcomes of this preliminary study can allow a wider range of researches in this direction. Indeed, if the ANNs are easier and faster to be trained, new simulations are possible and a whole new range of possibilities can become reality, such as training ANNs to understand criminal networks recurrent patterns.

1.3 Thesis Outline and Contributions

Finally, this section illustrates the thesis structure.

Chapt. 2 comprises all the background needed to understand the research herein presented. In particular, in Sect. 2.1 the theoretical definitions are provided whereas in Sect. 2.2 and in Sect. 2.3 the state-of-art about Sparse ANNs and Network Science are described, respectively. Lastly, Sect. 2.4 summarises of the relevant notions shown in the chapter. Chapters 3 - 6 are the experimental ones. Chapters 3 and 4 are strictly related to each other as they both analyse criminal networks in the contexts of: (i) disrupting the information spreading of criminal organisations, and (ii) detecting the most relevant effects of missing data on LEAs investigations. In particular, the work included in (i) was published in [16, 17] whereas the study conducted in (ii) has been published in [18]. Next, the network feature of resilience and, thus, the modelling of the probabilistic failure framework is proposed in Chapt. 5. This led to a paper that is undergoing peer-review in IEEE Transactions on Network Science and Engineering (accessible in ArXiv [19]). Finally, the last experimental chapter (Chapt. 6) covers the topic of the ANNs analysis in the specific context of Sparse ANNs, which has been published in [20, 21]. Lastly, in Chapt. 7, the conclusions are provided.

Chapter 2

Background

In this chapter, all the background needed to understand this thesis is provided. In particular, Sect. 2.1 collects all the theoretical materials and methods borrowed from Graph Theory and Network Science. Next, in Sect. 2.2 a focus on theory and state of art about Artificial Neural Networks is provided. Lastly, in Sect. 2.3 there is a more general overview of all the most significant works conducted so far by the academic community about Social Network Analysis, Criminal Networks with a focus on the Sicilian criminal organisation. Lastly, the studies about the Network robustness are shown.

2.1 Graph Theory

Graph Theory relates the study of graphs, that are mathematical structures used to model pairwise relations between objects (vertices connected by edges). An extension of this scientific field is Network Science that is a combination of several ones and among them there is Graph Theory itself. Indeed, Network Science draws Graph Theory from mathematics with the aim to study complex networks (*i.e.*, transportation networks, telecommunication networks and social networks). All those networks may be investigated through the use of Social Network Analysis (SNA) that is, thus, the process of investigating social structures through the use of Network Science and Graph Theory.

Thus, Graph Theory is a well established field in mathematics. However, only recently many of its theoretical results started to be used within Social Network Analysis, an area with significant implications for real–world scenarios. For ex-

ample, one can simulate the behaviour of social networks using strategies like link predictions [22, 23], temporal networks, or spreading of influences [24, 19]. Other practical applications include to deal with large Artificial Neural Networks [15, 20, 20], or ANN for short, or targeted advertisements to people based on their friends' interests [25], or containing the spread of fake news [26].

Network Science tools may also be used in the investigation of criminal networks. In the context of criminal organisations, for instance, the complex social interactions within a clan-based society may help the feature–selection process, as required for building machine learning models [27]. Other times, it is Network Science itself that helps conducting better-performing investigations by Law Enforcement Agencies (LEAs). To this end, criminal networks can be encoded as graphs, and various types of analysis and simulations can be carried out for modelling criminal behaviours.

This section is to be intended as a short tutorial on how Network Science strategies have been involved in the analyses we conducted and described in this thesis. It is divided into five parts: (i) in Sect. 2.1.1 the basic definitions of Graph Theory that relates the key theoretical tools are provided; (ii) Sect. 2.1.2 is a summary of the most important sparse networks properties that have a central role in understanding the experiments conducted in Chapt. 6; (iii) a description of the centrality metrics we used, especially in Chapters 3 and 5, can be find in Sect. 2.1.3; (iv) Sect. 2.1.4 relates an explanation of the distance metrics we selected for computing the similarity between graphs (Chapt. 4); (v) lastly, in Sect. 2.1.5 there is an overview on the evaluation metrics we chose to test the goodness of our approaches in some application domains (in Chapters 3 and 5).

2.1.1 Basic Definitions

In this section we introduce basic definitions which will be largely used throughout the thesis. For more details, see [28, 29].

Graph: A graph, G is a pair $G = \langle V, E \rangle$ in which V is the set of *vertices*, and $E = \{\langle i, j \rangle : i \in V \wedge j \in V\}$ is the set of *edges*. Note that in Network Science the vertices are called nodes (represented by the set of nodes N) and the edges are defined as links (represented by the set of links L); thus, those two pair of terms in this work will be used interchangeably. As an example, the Social Network *Facebook*® may be viewed as a graph, where the nodes represent users and links represent the friendship relationship among them. Is it also possible to define a

subgraph H of the graph G as a graph whose nodes and links are subsets of the nodes and links of G. Throughout this thesis, the number of vertices (resp., edges) in V (resp., E) are denoted by v (resp., $2m$). We say that G is *sparse* [30] (resp., *dense*) if $m \in O(n)$ (resp., $m \in O(n^2)$).

Directed and Undirected Graph: An *undirected graph* is a graph $G = \langle V, E \rangle$, where all the edges E between vertices V are bidirectional. An undirected graph is sometimes called an undirected network. In contrast, a graph where the edges point in a direction is called a *directed graph*. In this thesis, we dealt with *undirected graphs*; *i.e.*, if $\langle i, j \rangle \in E$ (where $\langle i, j \rangle$ represents the connection from the vertex v_i to v_j), then $\langle j, i \rangle$ belongs to E as well.

Weighted and Unweighted Graph: A *weighted graph*, denoted as G with a little abuse of notation, is a triplet $G = \langle V, E, W \rangle$ in which V is the set of vertices, E is the set of edges and $W : E \leftarrow \mathbb{R}^+$ is a function that maps an edge $\langle i, j \rangle$ onto a non-negative real number w_{ij}. If a graph is made by weights equals only to zero or one; *i.e.*, $w_{ij} = 1$, $w_{ij} = 0$, than it is called *unweighted graph*. In this thesis, we dealt with both.

Adjacency Matrix: A common way to represent relationships among nodes is the *adjacency matrix* $\mathbf{A}$. Indeed, any graph is associated with $\mathbf{A}$ that is a square matrix such that $\mathbf{A}_{ij}$ is equal to 1 if and only if there is an edge from node i to node j; 0, otherwise. The adjacency matrix of an undirected graph is *symmetric*; *i.e.*, a_{ij}=a_{ji}, all its eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are *real*, and the corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ will form an *orthonormal basis* in $\mathbf{R}^n$ [31]. Analogously, for weighted networks each weighted and undirected graph G is associated with a symmetric matrix $\mathbf{W}$ such that $\mathbf{W}_{ij} = w_{ij}$ if and only if there is an edge from i to j with weight w_{ij}, 0 otherwise. The largest eigenvalue λ_1 of $\mathbf{A}$ is also called *spectral radius*.

The adjacency matrix, along with the *Laplacian* and *Normalised Laplacian* matrices, are the most common representation matrices for a graph.

Degree of a node: The *degree* of a node n_i, denoted $deg(i)$ or k_i, is the number of incident links to n_i. The sum of the degrees of all nodes is equal to the double of the number of links L:

$$\sum_{n \in N} k_n = 2L.$$

In weighted networks, the *weighted degree* (also known as *strength* [32, 33]) is the sum of the edges weights w incident on n_i:

$$k_i = \sum_{(i,j)\in L} w_{ij},$$

where the summation spans over all links (i, j) in the network, linked to node n_i. For undirected networks, the *average degree* is defined as

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N},$$

where N is the total number of nodes (or vertices V), k_i is the degree of a generic node i, and L represents the total number of links (or edges E) within the network (or graph).

Degree and Weight Distribution: The *degree distribution* p_k provides the probability that a randomly selected node in the network has degree k. Since p_k is a probability, it must be normalised; *i.e.*,

$$\sum_{i=1}^{\infty} p_k = 1.$$

For a network made of N nodes, the degree distribution is the normalised histogram given by:

$$p_k = \frac{N_k}{N},$$

where N_k is the number of nodes having degree k.

The degree distribution has assumed a central role in network theory following the discovery of scale-free networks (See "Scale-Free Property" paragraph); moreover, p_k determines many network phenomena, from network robustness to the spread of viruses.

The degree distribution can be extended to weighted networks considering the weighted degree (strength) distribution $P(s)$, defined as the probability that a node may have weighted degree (strength) equal to s. Based on [33], this is

$$P(s) \sim s^{-\gamma},$$

where γ is a constant typical of the network.

Walks and Paths: A *walk* of length $r > 0$ is a sequence of alternating vertices and edges $i_1, e_1, i_2, e_2, \ldots, e_r, i_{r+1}$, such that for each $\ell = 1, \ldots, r$ the edge e_ℓ is $\langle i_\ell, i_{\ell+1} \rangle$. A *path* is a walk with no-repeated nodes. A graph is *connected* if every pair of nodes in it are connected through a path. A path is called *simple* if each vertex in the path is distinct. More formally, a path can be defined as a sequence of vertices

$$P = (v_1, v_2, \ldots, v_m) \in V \times V \times \cdots \times V,$$

such that v_i is adjacent to v_{i+1} for $1 \leq i \leq m-1$. Such a path P is called a path of length $m-1$ from v_1 to v_m. Measures based on paths strategies are the **shortest path length analysis**. The *distance* from a vertex v_i to a vertex v_j in G, denoted $d(v_i, v_j)$ is the length of a **shortest path** from v_i to v_j (if such a path exists).

$$d_{ij} = min\left(\Gamma(i,j)\right),$$

where $\Gamma(i,j)$ is the set of paths connecting i and j.

Thus, in a network, a *path* is a sequence of nodes such that each node is connected to the next one along the path by a link. Each path consists of $n+1$ nodes and n links. The length of a path is the number of its links, counting multiple links multiple times. It is a route that runs along the links of the network. The number of links the path contains is called *path length*. In weighted networks, the path's length is given by the sum of the weighted edges of the path. The *shortest path* from non-adjacent node n_i to n_j is the path with the fewest number of links. Multiple shortest paths of the same length d_{ij} can exist. The shortest path never contains loops or intersects itself. In an undirected network $d_{ij} = d_{ji}$. In directed networks, often $d_{ij} \neq d_{ji}$; indeed, in those kind of networks the existence of a path from n_i to n_j does not guarantee the existence of a path from n_j to n_i. Chapt. 3 reports our study on the shortest path lengths on our two criminal graphs to better show the communication behaviour inside a "cosca". This is an important tool to demonstrate how criminals act to avoid to overexpose their bosses using a balanced number of intermediates to safely communicate between mobster.

Connected Graphs: A graph G is *connected* if, for any two nodes, there is a path between them. If G is not connected, its maximal connected subgraphs are called the *connected components* of G. If a network consists of two components, a properly placed link can connect them, making the network connected. Such a link is called *bridge*.

Clustering Coefficient: Clustering is used to quantify the relationship among nodes' neighbours. Indeed, the degree only considers the number of direct links between nodes. The *clustering coefficient* C_i measures the link density in the immediate neighbourhood of a node. $C_i \in [0, 1]$ represents the clustering coefficient of a generic node n_i:

$$\begin{cases} \text{if } C_i = 0, & \text{there are no links among the node's neighbours} \\ \text{if } C_i = 1, & \text{each node's neighbour is connected with the others} \end{cases}$$

The local *Clustering Coefficient* (CC) is a measure of the degree to which nodes in a network tend to cluster together. For unweighted networks, the clustering of a node i, denoted by $CC(i)$, is the fraction of possible triangles through that node that exist, given by

$$\text{CC(i)} = \frac{2T(i)}{k_i(k_i - 1)},$$

where $T(i)$ is the number of triangles through node i and k_i is the degree of i. For weighted networks, the clustering is defined as the geometric average of the subgraph edge weights [34], obtained from the formula

$$\text{CC(i)} = \frac{1}{k_i(k_i - 1)} \sum_{i,j \in N} (\hat{w}_{ij} \hat{w}_{iw} \hat{w}_{jw})^{1/3},$$

where $\hat{w}_{ij}$ are the link weights normalised by the maximum weight in the network $\hat{w}_{ij} = w_{ij} / \max(w)$.

Average Clustering Coefficient: The average $\langle C \rangle$ of $C_i \in i = 1, \ldots, N$ in the whole network is given by

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i.$$

Scale-Free property: The majority of real networks, such as the World Wide Web, are called *scale-free networks* that means that their degree distribution follows a power law. The *power-law distribution* has the following form

$$p_k \sim k^{-\gamma},$$

where the exponent γ is its *degree exponent*. Some artificial network models (See Sect. 2.1.2) such as the **Barabási-Albert (BA) Model** successfully exhibit this feature.

Small-World phenomenon: The Small World phenomenon [35, 36] is based on the concept of the six degrees of separations, according to which two random people in the world may be connected each other via a few acquaintances (*i.e.*, it is estimated that there are six people in the middle between the source and the destination). In Network Science, it translates into a "short" distance between two randomly chosen nodes within a network, that is

$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}, \tag{2.1}$$

where N is the total number of nodes in the graph, $\langle k \rangle$ is the network average degree, and $\langle d \rangle$ the average distance within the network. The denominator implies that the denser the network, the smaller the distance between the nodes is. In conclusion, the average path length or the diameter depends logarithmically on the system size.

2.1.2 Sparse Networks

This section briefly explain the main characteristics of artificial networks. In particular, several configurations of random networks that introduce the concept of sparseness are described.

The need for scientists to create Artificial, or Synthetic Networks has been born from the aim to reproduce real network properties in a controlled environment. For this reason, several typologies of Artificial Networks have been formulated.

Random network theory emulates the irregularity and unpredictability of real networks by constructing from scratch and characterising graphs that are truly random. Some of the most popular random network models are *Erdös-Rényi* (ER) and its variation the *Gilbert Model*. Other well-known artificial network models are the *Watts-Strogatz* (WS) and *Barabási–Albert* (BA). In particular, this last one tries to capture two important properties of real network: the growth and the preferential attachment. Further details on those models are provided in the following paragraphs.

Random network: A random network consists of N nodes where each node pair $(n_i, n_j), \forall i, j \in N$ is connected with probability p. To construct a random network one needs to:

1. Start with N isolated nodes,
2. Select a node pair (n_i, n_j) and generate a random number $rand \in [0, 1]$:

$$\begin{cases} \text{if } rand > p, & \text{connect the selected node pair with a link} \\ \text{otherwise,} & \text{leave them disconnected} \end{cases}$$

3. Repeat the previous step for all pairs of distinct nodes $(n_i, n_j) \in N \times N$.

The network obtained after this procedure is called a *random graph* or a *random network*. There are two definitions of a random network: the definition provided in the Erdős-Rényi Model, and the one of the Gilbert Model.

***Erdős-Rényi* and *Gilbert* Models:** Random networks are also called *Erdős-Rényi Networks* from the names of the mathematicians Paul Erdős (1913-1996) and Alfréd Rényi (1921-1970), who studied the properties of these networks. According to the ER model [37] (*i.e.*, the $G(n, M)$ model) a network is firstly generated by laying down a number n of isolated nodes; next, the nodes are connected with M randomly placed links that is the model we used to conduct our experiments. Even though it is unlikely that real social networks form like this, such models can predict a number of different properties [37]. A closely related variant is the $G(n, p)$ model [38] defined by Edgar Nelson Gilbert (1923-2013) in which, after the initial generation of n isolated nodes, each pair is selected and a random number in the interval $[0, 1]$ is chosen. If such number exceeds a chosen probability p, then the selected nodes are connected. Otherwise, they are left disconnected. The procedure is performed for all the $n(n-1)/2$ pairs of nodes.

There are two main limits in Random Network Model that had to be overcome over the years by the academic community:

1. The local clustering coefficient in ER model is given by [28]

$$C_i = \frac{\langle k \rangle}{N}.$$

This behaviour of C_i is contradicted by the local clustering coefficient of real networks.

2. The Poisson distribution that describes the degree distribution of ER networks does not allow large differences between the worst- and best-connected nodes in the network. This implies that hubs, frequently observed in real networks, cannot be found in ER networks. BA model, relying on preferential attachment and growth, successfully reproduces this fundamental feature.

***Watts-Strogatz* Model:** While the ER model may exhibit a small clustering coefficient along with a small average shortest path length, the WS model [39] can produce graphs with *small-world* properties that are highly clustered, but with small characteristic path lengths. Most nodes are not neighbours, but the neighbours of a node are likely to be connected and most nodes can be reached from every other one by a small number of steps (also called *Six Degree of Separation* property) [39]. Indeed, the two main considerations motivated Duncan J. Watts (1971) and Steven Strogatz (1959) to propose this model are that (i) in real networks the average distance between two nodes depends logarithmically on N (ii) the average clustering coefficient $\langle C \rangle$ of real networks is much higher than expected for a random network of similar N and L. As explained in the previous section, in a small-world network, if d is the distance in steps between two randomly chosen nodes, it grows proportionally to the logarithm of the number of nodes n: $d \propto \log(n)$. Thus, to construct the model is needed to start from a ring of nodes, each node is connected to their previous and next neighbours. Each link is then rewired with probability p to a randomly chosen node. For small values of p, the network maintains high clustering, but the random long-range links can drastically decrease the distances between the nodes. When $p = 1$, all links are rewired, so the network turns into a random ER network [39]:

$$\begin{cases} \text{if } p \simeq 0, & \text{regular lattice} \\ \text{if } 0 < p < 1, & \text{Small-World property} \\ \text{if } p = 1, & \text{Random Network Model (all links rewired).} \end{cases}$$

The Watts-Strogatz model interpolates between a *regular lattice*, which has high clustering (but lacks the Small-World phenomenon), and a *random network*, which has low clustering (but displays the Small-World property). Moreover, high nodes degrees are absent from Watts-Strogatz model.

***Barabási-Albert* Model:** This model was theorised by Albert-László Barabási (1967) and Réka Albert (1972) [40]. The BA model [40] exploits a preferential

attachment mechanism to develop a *scale-free network*; *i.e.*, the degree distribution follows a power law. The algorithm starts from a network with m_0 nodes, whose links are chosen arbitrarily, as long as each node has at least one link. At each step, a new node with $m \leq m_0$ links is added. The preferential attachment ensures that the probability p_i that the new node is connected to a node i depends on the degree d_i of the latter as follows:

$$p_i = \frac{d_i}{\sum_j d_j}.$$

Thus, the new node prefers to attach itself to already heavily linked nodes, called *hubs*, which tend to accumulate even more links at each step, while nodes with only few links are unlikely to be chosen [40].

To summarise an artificial network with the Barabási-Albert model, the steps are:

1. Start with a set of N_0 nodes, the links between that are chosen arbitrarily, as long as each node has at least one link.
2. **Growth** – At each time-step a new node n_j with l links (with $l \leq l_0$) that connects the new node to nodes already in the network is added.
3. The connections between the new node with the older nodes are defined by the **Preferential Attachment** probability.

2.1.3 Centrality Metrics

Centrality is a key concept in network analysis, and refers to the importance of a node in a network. The notion of node centrality has been introduced in late 1940s to quantify the importance of an actor in a social network [30]. Roughly speaking, a *centrality metric* is a function $\phi : N \to \mathbb{R}^+$ that maps a node i of a graph onto a real non-negative number $\phi(i)$, under the assumption that the larger $\phi(i)$, the more important i. There are multiple measures for network centrality in use within SNA. However, in this section there is a description of the most popular ones, that we also considered in conducting our experiments; *i.e.*, (i) Degree, (ii) H-Index, (iii) Coreness, (iv) Betweenness, (v) Eigenvector, (vi) Katz, (vii) Collective Influence. In particular, the most effective ones in detecting, for instance, the strategic positions in criminal networks are Degree centrality and Betweenness centrality [41, 42].

Some centrality metrics such as *degree*, *h-index*, and *coreness* depend on the ability

of a node to influence its surrounding neighbours; hence, the first centrality to be described below is the *degree* - this is the most straightforward node centrality metric.

Degree Centrality: This [43] is a measure that evaluates the local importance of a node within the graph; given a node i, the Degree centrality $C_D(i)$ of i is defined by:

$$\mathrm{C_D(i)} = \sum_{j \in N, j \neq i} a_{ij},$$

where $A = (a_{ij})$ is the adjacency matrix of the graph.

The *node degree* represents the number of links adjacent to the node, while the *weighted node degree* is the sum of the link weights, for links incident to that node. This measure has been formalised as follows [44]:

$$\mathrm{C_D^W(i)} = \sum_{j \in N, (i,j) \in L} w_{ij},$$

where w is the weighted adjacency matrix, in which w_{ij} is greater than 0 if the node i is connected to node j, and the value represents the weight of the link.

In other words, the degree centrality can be defined as:

$$\mathbf{d} = \mathbf{A} \times \mathbf{1}$$

where $\mathbf{d}$ is a vector whose i-th entry denotes the degree of node i, $\mathbf{1}$ is the vector whose entries are all equal to 1, and the symbol $\times$ denotes the usual matrix-by-matrix (or matrix-by-vector) product.

H-Index Centrality: The h-index, or *Hirsch index* (as the name of its creator), is a parameter that quantifies the academic impact of scientists [45]. For our purposes, the h-index is a local centrality metric defined as follows: a node i in a graph G has h-index h if i has at least h neighbours, each of them with degree greater than or equal to h.

Coreness Centrality: The *coreness* [46] of a node is grounded on the computation of the ζ-*core* $G_\zeta = \langle N_\zeta, L_\zeta \rangle$ of a graph $G = \langle N, L \rangle$, being ζ a positive integer. Formally, the graph G_ζ is a subgraph of G, which satisfies the following properties: (a) each node in N_ζ has degree of at least ζ, and (b) the graph G_ζ is *maximal* against property (a); *i.e.*, if the aim is to add any node $j \in N - N_\zeta$ to G_ζ

along with its links, then the property (a) would no longer hold true. Based on this definition, it may be asserted that a node $i \in N$ has coreness ζ if it belongs to G_ζ, and it does not belong to $G_{\zeta+1}$. Both the h-index and the coreness of a node are clearly related to the degree of that node, as reported in [46].

Betweenness Centrality: Some nodes may play a particularly important role in propagating information because they act as bridges between separate regions of a graph and so they have the potential to slow down (or magnify) the information flow from one region to another. Such nodes are said to have a high value of *Betweenness centrality* [47]. Specifically, the (shortest-path) betweenness $C_B(i)$ of a node i is defined as follows:

$$\mathrm{C_B(i)} = \sum_{s,t \in N, s \neq t} \frac{\sigma(s,t \mid i)}{\sigma(s,t)},$$

where $\sigma(s,t)$ is the number of shortest paths between an arbitrary pair of nodes s and t, while $\sigma(s,t \mid i)$ denotes those shortest paths passing through the node i.

Betweenness centrality is a measure based on shortest paths for both unweighted and weighted networks. For our algorithm, we use Breadth-First Search (BFS) for unweighted and Dijkstra's algorithm for weighted graphs.

Eigenvector Centrality: There are some metrics, such as the *Eigenvector* centrality [48] and the *Katz* centrality [49] (that will be described afterwords) that rely on the full knowledge of the graph topology. Given a constant $\lambda \neq 0$, the Eigenvector centrality $\mathbf{e}$ is defined as the solution to the following equation:

$$\mathbf{Ae} = \lambda \mathbf{e}$$

where $\mathbf{e}$ is a vector whose i-th entry denotes the Eigenvector centrality of the node i. If G is connected and $\lambda = \lambda_1$, then the Perron-Frobenius Theorem [30] states that there is a unique solution with all components positive to the equation just stated that corresponds to the largest eigenvector of $\mathbf{A}$.

Katz Centrality: *Katz centrality* [49] is another centrality measure, which defines the centrality for a node based on the centrality of its neighbours. For a node i this is defined as:

$$\mathrm{C_K(i)} = \alpha \sum_{j \in V} a_{ij} \mathrm{C_K(j)} + \beta,$$

where α and β are positive constants, $A = (a_{ij})$ is the adjacency matrix of the graph, whose eigenvalues are denoted by λ_i, $i = 1, \ldots, n$. The parameter β controls the initial centrality, while the parameter α satisfies the inequality:

$$\alpha < \frac{1}{\max\{\lambda_i : 1 \leq i \leq n\}}.$$

The Katz centrality for weighted networks can be computed in a similar way, but in this case we have to use the weighted adjacency matrix instead.

In other words, Katz centrality can also be defined as $\mathbf{k}$:

$$\mathbf{k} = (\mathbf{I} - \alpha\mathbf{A})^{-1} \times \mathbf{1}$$

where, $\mathbf{I}$ is the identity matrix and α is a fixed parameter that must be smaller than $1/\lambda_1$. If $\alpha \simeq 0$, then Katz centrality well approximates the degree. On the other hand, if $\alpha \simeq \frac{1}{\lambda_1}$, then Katz centrality is a good approximation of the Eigenvector centrality [50, 51]. The semantics of the Katz centrality is as follows: given a node i, let us consider all walks of arbitrary length starting from i and ending in any other node j. Node i is assumed to be important if it is well connected to any other node through walks of arbitrary length. Yet, shorter walks have to be preferred to longer ones. To this purpose, walk length is weighted through a decreasing factor α.

Collective Influence Centrality: Another useful network metric is the *Collective Influence* (CI) [52], which computes the centrality (or influence) of a node i of a network according to the formula:

$$\mathrm{CI}_\ell(\mathrm{i}) = (k_i - 1) \sum_{j \in \delta B(i,\ell)} (k_j - 1),$$

where k_i is the degree of node i, $B(i, \ell)$ is the ball of radius ℓ centred on node i, and $\delta B(i, \ell)$ is the frontier of the ball, that is, the set of nodes at distance ℓ from i (the distance between two nodes is defined as the number of links of the shortest path connecting them). To compute $CI_\ell(i)$, we first find the nodes on the frontier $\delta B(i, \ell)$. To compute the CI in a weighted network, we have to substitute the degree k of a node by his weighted degree $\mathrm{C}_\mathrm{D}^\mathrm{W}(\mathrm{i})$.

2.1.4 Distance Metrics

In this section we describe the distance metrics adopted hereafter to compare the differences between two graphs. In particular, the spectral and matricial distances are shown.

Spectral Distances: As already mentioned in the previous sections of this chapter, the *spectrum* of a graph consists of the set of sorted (increasing or decreasing) eigenvalues of one of its representation matrices. It is used to characterise graph properties and extract information from its structure. The spectra derived from each representation matrix may reveal different properties of the graph. The largest eigenvalue (in absolute value) of the graph is called the graph's *spectral radius*. In the case of the adjacency matrix A, if λ_k is its k^{th} eigenvalue, the spectrum is given by their descending order as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

Spectral distances allow to measure the structural similarity between two graphs starting from their spectra.

The most common matrix representations of a graph are the *Adjacency matrix* A, the *Laplacian matrix* L, and the *normalised Laplacian* $\mathcal{L}$.

In order to explain those kind of spectral distance metrics, we will firstly have to describe how the Laplacian and the Normalised Laplacians matrices are made. Then, we will show their spectral structures. Lastly, we will enunciate the formula of those three spectrum distances (*i.e.*, Adjacency, Laplacian and Normalised Laplacian).

Given a graph G with n nodes, its adjacency matrix A, as already shown in the previous section, is an $n \times n$ square matrix denoted by $A = (a_{ij})$, with $1 \leq i, j \leq n$, where $a_{ij} = 1$ if there exists an edge joining nodes i and j, and $a_{ij} = 0$ otherwise. For undirected graphs the adjacency matrix is symmetric, *i.e.*, a_{ij}=a_{ji}.

The degree matrix D is a diagonal matrix where $D_{ii} = k_i$ and $D_{ij} = 0$ for $i \neq j$.

$$D_{ij} = \begin{cases} k_i & \text{if} i = j \\ 0 & \text{otherwise.} \end{cases}$$

The adjacency matrix and the degree matrix are used to compute the combinatorial Laplacian matrix L, which is an $n \times n$ symmetric matrix defined as

$$L = D - A.$$

The diagonal elements L_{ii} of matrix L are then equal to the degree k_i of the node i, while the off-diagonal elements L_{ij} are -1 if the node i is adjacent to j, and 0 otherwise. A normalised version of the Laplacian matrix $\mathcal{L}$ is defined as

$$\mathcal{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}},$$

where the diagonal matrix $D^{-\frac{1}{2}}$ is given by

$$D_{i,i}^{-\frac{1}{2}} = \begin{cases} \frac{1}{\sqrt{k_i}} & \text{if } k_i \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

If the representation matrix is symmetric, its eigenvalues are real and they can be sorted. The spectrum of a graph consists, indeed, of the set of the sorted eigenvalues of one of its representation matrices. The sequence of eigenvalues may be ascending or descending depending on the chosen matrix. The spectra derived from each representation matrix may reveal different properties of the graph. The largest eigenvalue in modulus is called the *spectral radius* of the graph. If λ_k^A is the k^{th} eigenvalue of the adjacency matrix A, then the spectrum is given by the descending sequence

$$\lambda_1^A \geq \lambda_2^A \geq \cdots \geq \lambda_n^A.$$

If λ_k^L is the k^{th} eigenvalue of the Laplacian matrix L, such eigenvalues are considered in ascending order so that

$$0 = \lambda_1^L \leq \lambda_2^L \leq \cdots \leq \lambda_n^L.$$

The second smallest eigenvalue of the Laplacian matrix of a graph is called its *algebraic connectivity*. Similarly, if we denote the k^{th} eigenvalue of the normalised Laplacian matrix $\mathcal{L}$ as $\lambda_k^{\mathcal{L}}$, then its spectrum is given by

$$0 = \lambda_1^{\mathcal{L}} \leq \lambda_2^{\mathcal{L}} \leq \cdots \leq \lambda_n^{\mathcal{L}}.$$

Then, the *spectral distance* between two graphs is the euclidean distance between their spectra [53]. Given two graphs G and G' of size n, with their spectra respectively given by the set of eigenvalues λ_i and λ'_i, their spectral distance, according

to the chosen representation matrix, is computed as follows by the formula

$$d(G, G') = \sqrt{\sum_{i=1}^{n} (\lambda_i - \lambda'_i)^2}.$$

Hence, we obtained the adjacency spectral distance d_A, the Laplacian spectral distance d_L and the normalised Laplacian spectral distance $d_{\mathcal{L}}$.

If the two spectra are of different sizes, the smaller graph is brought to the same cardinality of the other by adding zero values to its spectrum. In such case, only the first $k \ll n$ eigenvalues are compared. Given the definitions of spectra of the different matrices, the adjacency spectral distance d_A compares the largest k eigenvalues, while d_L and $d_{\mathcal{L}}$ compare the smallest k eigenvalues. This determines the scale at which the graphs are studied, since comparing the higher eigenvalues allows to focus more on global features, while the other two allow to focus more on local features.

Matricial Distances: Another class of distances between graphs is the matrix distance [54]. A matrix of pairwise distances d_{ij} between nodes on the single graph is constructed for each as

$$M_{ij} = d_{ij}.$$

While the most common distance d is the shortest path, other measures can also be used, such as the effective graph resistance, or variations on random-walk distances. Such matrices provide a signature of the graph characteristics and carry important structural information. Matrices M are then compared using some norm or distance.

Given two graphs G and G', having M and M' as their respective matrices of pairwise distances, the matrix distance between the G and G' is introduced as:

$$d(G, G') = \|M - M'\|,$$

where $\|.\|$ is a norm to be chosen. If the matrix used is the adjacency matrix A, the resulting distance is called *edit distance*.

The similarity measure used in this work is called DeltaCon [3]. It is based on the root euclidean distance $\mathrm{d}_{\mathrm{rootED}}$, also called *Matsusita difference*, between matrices S created from the fast belief propagation method of measuring node

affinities.

The DeltaCon similarity sim_{DC} is defined as

$$sim_{DC}(G, G') = \frac{1}{1 + \mathrm{d}_{\mathrm{rootED}}(G, G')},$$

where the root euclidean distance $\mathrm{d}_{\mathrm{rootED}}(G, G')$ is defined as

$$\mathrm{d}_{\mathrm{rootED}}(G, G') = \sqrt{\sum_{i,j} (\sqrt{S_{i,j}} - \sqrt{S'_{i,j}})^2}.$$

When used instead of the Euclidean distance, $\mathrm{d}_{\mathrm{rootED}}(G, G')$ may even detect small changes in the graphs. The fast belief propagation matrix S is defined as

$$S = [I + \varepsilon^2 D - \varepsilon A]^{-1},$$

where $\varepsilon = 1/(1 + \max_{1 \leq i \leq n} D_{ii})$ and it is assumed to be $\varepsilon \ll 1$, so that S can be rewritten in a matrix power series as:

$$S \approx I + \varepsilon A + \varepsilon^2 (A^2 - D) + \dots.$$

Fast belief propagation is an effective algorithm and it is designed to perceive both global and local structures of the graph [3].

2.1.5 Evaluation Metrics

In this section we describe the evaluation metrics adopted hereafter to verify the results obtained from our experiments. In particular, the Spectral Radius, the Largest Connected Component (LCC) and the Average Path Length (APL) are shown.

The Spectral Radius of a graph and its role in governing dynamic processes: The largest eigenvalue λ_1 of the adjacency matrix of a graph G – also known as the *spectral radius* – can be used to analyse dynamical processes taking place over G, such as: the spread of a flu-like epidemics over a population or the spread of a malware in a computer network [55, 7, 11].

Early studies on virus propagation in human population pointed out the existence of a threshold R_0 (called *virus reproduction number*) such that if $\lambda_1 \geq R_0$ then a

virus causes a global pandemics; whereas if $\lambda_1 < R_0$ the virus gets wiped out [56, 7, 8].

Due to its practical relevance, many authors were interested in assessing how λ_1 varies upon the removal of a target node [57, 58]. Specifically, Tong *et al.* [58] introduced the *k-node deletion problem*, which can be stated as follows: given an undirected graph $G = \langle N, E \rangle$, find the set of nodes $\mathcal{S}^\star(k) \subseteq N$ of cardinality k which, if deleted from G, yield the biggest drop in λ_1. The k-node deletion problem is NP-Hard [58], thus efficient but accurate approximation algorithms are required to solve it. The state-of-the-art solution to the k-node deletion problem is the *NetShield* algorithm [55], which achieves a worst-case time complexity of $O(nk^2 + m)$, being n and m the number of nodes and edges in G, respectively.

The largest connected component of a graph: In graph theory, the LCC is a way to measure the network's connectivity. Its size defines how many people a single individual (*i.e.*, a node) is able to reach through its relationship bonds (*i.e.*, direct links/edges or paths). Indeed, if the LCC size has the same order of magnitude of the network size, then the connectivity is at its maximum. On the other hand, nodes removal (jointly with their links) may provoke the emerging of smaller clusters (or even isolated nodes). As a consequence, a LCC size drop implies that the network becomes less and less connected. Thus, the LCC size allows to quantify the effectiveness of the nodes removal strategies.

Early studies investigated the decrease in network connectivity due to the selective removal (also known as *attack*) of some of the network nodes or edges [40, 59, 60, 5]. An interesting class of attacks consists of repeatedly increasing the number of nodes/edges deleted from a graph G. This operation implies that G breaks into disconnected subgraphs; thus, an important parameter to assess the ability of G to preserve its functionality is given by the size c of its largest connected component (LCC); *i.e.*, the largest connected subgraph in G after node/edge removal.

Studies in the field of OSNs indicate that c is in the same order of magnitude of the the entire network; thus, the LCC is also called *giant component* [59, 61]. Studies on the LCC size are also closely related to the topic of *percolation* and to the structure of random graphs [30]. For instance, if an Erdős-Rényi random graph [37] of n nodes is considered, in which edges are placed uniformly at random between pair of nodes with probability p_e, then [61] proved that there exists a constant Ψ such that if $p_e \geq p_e^\star = \frac{(1+\Psi)}{n}$; thus, there exists a giant component in G containing $O\left(n^{\frac{2}{3}}\right)$ nodes. On the contrary, if $p_e < p_e^\star$, then all the connected components of G have the average size of $O(\log n)$.

We define the *transient phase* as the step in which G moves from a highly-connected state to a new one in which the removal of a sufficiently high number of nodes leads to a significant decrease in the LCC size. The fragmentation process deriving from node removal is not gradual: it is characterised by a critical threshold f_c. If the fraction f of removed nodes is less than f_c, then a giant component persists; but, once $f \geq f_c$, the giant component vanishes [30, 61].

The *Average Path Length*: This is another useful metric [62] for assessing the connectivity of a graph. It is the average number of steps along the shortest paths for all possible pairs of network nodes. The average path length is defined as follows:

$$\text{APL} = \sum_{i,j \in V} \frac{d(i,j)}{n(n-1)} \tag{2.2}$$

where $V = \{1, \ldots, n\}$ is the set of nodes (or vertices) in the graph G, $d(i,j)$ is the length of the shortest path from node i to node j, and n is the number of nodes in G.

2.2 Sparse Artificial Neural Networks

This section provides a brief overview of theory and methods used to induce sparse Artificial Neural Networks (ANNs), which have been used for conducting the experiments in Chapt. 6. In order to do so, Sect. 2.2.1 describes the theoretical concept behind our study. Next, the existing methods have been classified in two main categories, namely: (i) Methods derived from Network Science to induce sparse ANNs (Sect. 2.2.2), (ii) Methods derived from ANN Regularisation to induce sparse ANNs (Sect. 2.2.3).

2.2.1 Theory behind Sparse Artificial Neural Networks

This section briefly introduces the main concepts required for understanding a piece of work of this thesis. Note that, for the sake of simplicity, the words "weight" and "link" are used interchangeably, and only weighted links have been considered. The goal is to demonstrate the effectiveness of the SET framework [15], aiming at lower revise fraction values, in the context of the multilayer perceptron (MLP) supervised model. MLP is a feed-forward ANN composed by several hidden layers, forming a Deep Network, as shown in Fig. 2.1. Because of the intra-layer links

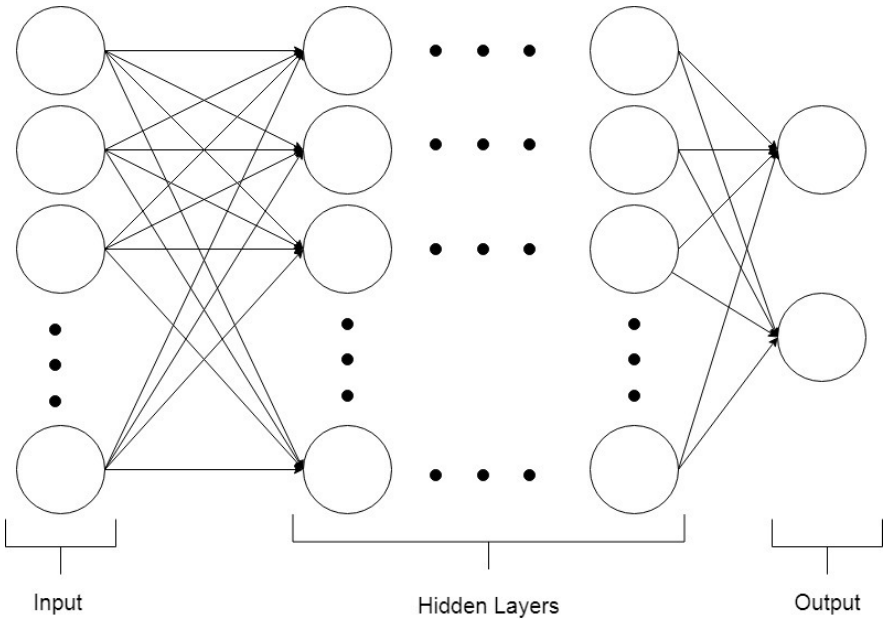


Figure 2.1: Example of a generic multilayer perceptron network with more than two hidden layers. Circles represent neurons, and arrows describe the links between layers.

flow, an MLP can be seen as a fully connected directed graph between the input and output layers.

Supervised learning involves observing several samples of a given dataset, which will be divided into "training" and "test" samples. While the former is used to train the neural network, the latter works as a litmus test, as it is compared with the ANN predictions. One can find further details on Deep Learning in [63, 64].

The construction of a fully connected graph inevitably leads to higher computational costs, as the network grows. To overcome this issue, the SET framework [15], drew inspiration from human brain models and modelled an ANN topology as a weighted sparse Erdős-Rényi graph in which edges were randomly placed with nodes, according to a fixed probability [37, 28, 29].

Like in [15], the edge probability is defined as follows:

$$p\left(W_{ij}^{k}\right) = \frac{\epsilon(n^{k} + n^{k-1})}{n^{k} n^{k-1}}, \tag{2.3}$$

where $W^{k} \in R^{n^{k-1} \times n^{k}}$ is a sparse weight matrix between the k-th layer and the previous one, $\epsilon \in R^{+}$ is the sparsity parameter, and i, j are a pair of neurons; moreover, n^{k} is the number of neurons in the k-th layer.

As outlined in the previous section, this process led to forcing network sparsity.

This stratagem is balanced by introducing the tunable revise fraction parameter ζ, which defines the weights fraction size that needs to be rewired (with a new weight assignment) during the training process.

Indeed, at the end of each epoch, there is a weight adjustment phase. It consists of removing the closest-to-zero links in between layers plus a wider revising range (*i.e.*, ζ). This parameter verifies the correctness of the forced-to-be-zero weights. Subsequently, the framework adds new weights randomly to exactly compensate the removed ones. Thanks to this procedure, the number of links between layers remains constant across different epochs, without isolated neurons [15].

Thus, in Chapt. 6 the role of ζ is analysed as well as showing how to find a good range of ζ values. Our aim is to to strike a good balance between learning speed and accuracy.

2.2.2 Methods derived from Network Science to induce sparse ANNs

Some previous papers focus on the interplay between Network Science and Artificial Networks [65, 15, 66]. More specifically, they draw inspiration from biological phenomena such as the organisation of human brain [29].

Early studies in Network Science, in fact, pointed out that real graphs (e.g., social networks describing social ties among members of a community) display important features such as power-law distribution in node degree [28] and the small-world property [39]. Many authors agree that these properties are likely to exist in many large networked systems one can observe in nature. For instance, in case of biological and neuronal networks, [67] suggested that the neuronal network describing the human brain can be depicted as a globally sparse network with a modular structure.

As a consequence, approaches based on Network Science consider ANNs as sparse networks whose topological features resemble those of many biological systems and they take advantage from their sparseness to speed up the training stage.

A special mention goes to recent research in [68], where the authors managed to train a million-node ANN on non-specialised laptops, based on the SET framework that was initially introduce in [15]. SET is a training procedure in which connections are pruned on the basis of their magnitude, while other connections are randomly added. The SET algorithm is actually capable of generating ANNs

that have sparsely connected layers and, yet, achieve excellent predictive accuracy on real datasets.

Inspired by studies on rewiring in human brain, [69] formulated the DEEPR algorithm for training ANNs under connectivity constraints. DEEPR automatically rewires an ANN during the training stage and, to perform such a task, it combines a stochastic gradient descent algorithm with a random walk in the space of parameters to learn.

Bourely et al. [66] studied to what extent the accuracy of an ANN depends on the density of connections between two consecutive layers. In their approach they proposed Sparse Neural Network architectures, which derive from random or structured bipartite graphs. Experimental results show that, with a properly chosen topology, Sparse Neural Networks can equal or supersede a fully connected ANN with the same number of nodes and layers in accuracy, with the clear advantage of handling a much smaller parameter space.

Stier et al. [65] illustrated a procedure to generate ANNs, which derives from artificial graphs. The proposed approach generates a random directed and acyclic graph G according to the Watts-Strogatz [39] or the Barabási-Albert [28] models. Nodes in G are then mapped onto layers in an ANN, and some classifiers (such as Support Vector Machines and Random Forest) are trained to decide if a Watts-Strogatz topology yields a better accuracy than a Barabási-Albert one (or *vice versa*).

2.2.3 Methods derived from ANN Regularisation to induce sparse ANNs

Methods such as L_1 or L_0 regularisation, which gained popularity in supervised learning, have been extensively applied to generate compact yet accurate ANNs.

For instance, [70] introduced additional gate variables to efficiently perform model selection. Furthermore, [71] described an L_0-norm regularisation method, which forces connection weights to become zero. Zero-weight connections are thus pruned, and this is equivalent to induce sparse networks.

The methods above are successful in producing sparse but accurate ANNs; however, they lack explainability. Thus, it is hard to understand why certain architectures are more competitive than others.

It is also interesting to point out that regularisation techniques can be viewed as

procedures compressing an ANN by deleting unnecessary connections (or, in an equivalent fashion, to select only few parameters). According to [72], techniques to prune an ANN are effective to uncover sub-networks within an ANN whose initialisation made the training process more effective. According to these premises, Frankle and Carbin suggested what they called the *lottery ticket hypothesis*. In other words, dense and randomly-initialised ANNs contain sub-networks (called *winning tickets*) that, when trained in isolation, are able to reach the same (or a comparable) test accuracy as the original network, and within a similar number of iterations.

2.3 Network Science Literature

In this section we provide an overview of the state-of-art of (i) Criminal Networks; (ii) Social Network Analysis applied to them and to (iii) the specific context of Sicilian Criminal organisations; lastly, (iv) the resilient properties of networks are considered.

2.3.1 Criminal Networks

Criminal organisations can be defined as groups operating outside the boundaries of the law, that profit from providing illicit goods and services in public demand in an illicit manner, and for which achievements come at the detriment of other individuals, groups or societies [73]. Organised crime can be referred to by a range of different terms such as *gangs* [74], *crews* [75], *firms* [76], *syndicates* [76], or *Mafia* [1]. In particular, Mafia is defined in Gambetta's work [77] as a "territorially based criminal organisation that attempts to govern territories and markets" and he refers to the one located in Sicily as the *original Mafia*.

Whatever term is used to identify organised crime, the latter is anyway based on relational traits. For this reason, scholars and practitioners are increasingly adopting a Social Network Analysis (SNA) perspective to explore criminal phenomena [78]. SNA is indeed a powerful tool to analyse criminal networks and to gain a deeper understanding of criminal behaviour [79]. SNA algorithms are able to produce relevant measurements and parameters relevant to identify the roles and importance of individuals within criminal organisations [80] and to construct crime prevention systems [81]. In a criminal investigation, the individuals subjected to Law Enforcement Agencies (LEAs) enquiries may attempt to shield sensible information. Investigators then have to rely on alternative methods and

exercise special investigative powers allowing them to gather evidence covertly. Information available for analysis can then come from sources such as phone taps, surveillance, archives, informants, interrogations to witnesses and suspects, infiltration in criminal groups. Despite significant advantages, such sources may also come with a number of drawbacks.

During investigations, some of the individuals providing information might be reliable, while others might attempt to deceive the investigations with the aim to protect themselves or their associates, or to achieve a goal. For instance, if actors are aware of being phone-tapped, they are more likely to avoid exposing some self-incriminating evidence. While transcripts of discussions between unsuspecting actors may be considered more reliable, a double-check is still needed between information collected from the taps and data collected from other official records related to the case. This is required since conversations among criminals often involve lies or codes concealing the true nature of the crime [82]. Moreover, if police misses surveillance targets, central actors may not appear with their actual role in the data, simply because their phones end up not being tapped [1].

While the police seeks to validate the content of phone-taps, the offenders themselves try to find out whether the information received from fellow criminals is actually accurate. Longer investigations and surveillance tend to eventually expose such lies. On the other side, with investigations going on, the list of suspects may change over time, with the group, and consequently data, changing significantly as a function of police decisions.

Police decisions may indeed impact the design of an investigation. LEAs normally start with some central individuals and then expand their reach by adding further actors. Not all the individuals linked to the central ones are automatically added, though. This can happen, for instance, when there are not enough resources available to investigate all active criminal groups, then prosecution services concentrate indeed on groups on which they can gather evidence easily. This kind of decisions are more prone to the risk of some groups operating under the police radar and then left out from the collected data. This approach is shown to hold extremely high chances of generating distorted inferences about the network structure [83].

The problem of actors lying is extended to data collected through questionnaires or interviews as well. Information collected from interrogations may not be reliable, with the risk of interviewees downplaying or amplifying their real role or not being representative of the broader group.

Incompleteness and incorrectness in criminal network data is then inevitable, since available intelligence data is determined more by the subjective judgements of investigators. This is due to investigators dealing with different qualities of data and because there is no standard methodology in SNA, for taking into account such degrees of reliability.

The problem of determining which information is relevant is usually referred to as the problem of *signal and noise*, in which important information is mixed in with large amounts of irrelevant, or unreliable information. LEAs are indeed often faced with the problem of having too much data, some of which being of little value. With large volumes of raw data collected from multiple sources, the risk of inconsistency becomes higher as well. Analytic techniques used in intelligence then must be able to cope with large amounts of information, and be capable to extract the signal from the noise.

In summary, data collected in criminal investigations often suffers from:

- *Incompleteness*, caused by the covert nature of such type of networks;
- *Incorrectness*, caused by either unintentional data collection errors and intentional deception by criminals;
- *Inconsistency*, when records of the same actors may be collected into law enforcement databases multiple times and not necessarily in a consistent way. Such misleading information may lead to an actor featuring multiple times (as different individuals) in the network.

Another problem specific to SNA for criminal networks lies in how data are transformed. As stated before, data needs to be presented in a specific manner, with actors being represented by nodes, whereas their associations or interactions are represented by links. In SNA, there is not a standard method for such data transformation task from raw data: the process undergoes the subjective judgement of the analyst that might be debatable. For instance, it may be difficult for an analyst to decide whom to include or exclude from the network, if its boundaries are prone to ambiguity [41]. Data conversion then ends up being a fairly labor-intensive and time-consuming task.

Finally, another feature of criminal networks is represented by their dynamics: such networks are not static, meaning that they constantly change over time. To represent such dynamics, new data or even different data collection methods are required, for covering longer time spans [41].

2.3.2 Social Network Analysis in Criminal Networks

Social Network Analysis is the use of Network and Graph Theory to study social phenomena, which was found to be highly relevant in areas like Criminology. This section provides an overview of key methods and tools that may be used for the analysis of criminal networks, which are presented in a real–world case study.

Social Network Analysis (SNA) is increasingly used by Law Enforcement Agencies (LEAs) to analyse criminal networks as well as to investigate the relations among criminals based on calls, meetings and other events derived from investigations [41, 79, 84, 85], but also to evaluate the effectiveness of law enforcement interventions aimed at disrupting criminal networks [86].

For this reason, nowadays there is a growing interest in the application of Graph and Network Science onto criminal networks. For instance, SNA has been used in [87] to build crime prevention systems. However, due to the lack of data availability (see Chapt. 4) on those kind of networks, there are difficulties in finding relevant quantitative studies. Such examples are those conducted by Szymanski [88] and Berlusconi [89], on the problem of community detection and link prediction.

Sarnecki [90] applied Social Network Analysis to study co-offending behaviours among Swedish teenagers. Morselli [91] studied the connections within the Gambino, New York based family, and he focused on the career of Saul Gravano, a member of the Gambino family. McGloin [92] made an analysis on the structure of a network based on the street gangs in Newark, New Jersey. Natarajan [93], built a network of phone calls starting from a dataset consisting of 2,408 wiretap conversations (gathered during the prosecution of a heroin-dealing Mafia syndicate in New York). This network revealed the core of the criminal organisation and showed that most of the members had very limited contacts with others in the group. Calderoni [94] explained how illicit drug traffics were indirectly handled by high-status Mafia members, whereas the most central and visible positions were held by middle-level criminals.

SNA is not only a tool to describe the structure and functioning of a criminal organisation, but it is largely employed in the construction of crime prevention systems [87]. For instance, Xu and Chen [84] jointly applied SNA, using hierarchical clustering algorithms. Their approach worked in two stages: firstly, a criminal network was partitioned into subgroups using a clustering algorithm; secondly, block modelling techniques have been used to extract interaction patterns between these subgroups. Agreste *et al.* [60] applied percolation theory to efficiently dismantle

mafia syndicates. Calderoni and Superchi [95] showed that the node's betweenness centrality in a meetings network is evidence of Mafia leadership, suggesting that this variable could be exploited by LEAs in selecting the most suitable targets for additional investigations and disruption. Social Network Analysis tools were also used to identify leaders within a criminal organisation. For instance, Mastrobuoni and Patacchini [96] investigated the structure of criminal ties between mobsters using a dataset of 800 Mafia members' criminal profiles. These criminals were active in the United States from 1950s to 1960s. Authors considered various features (such as family relationships, legal and illegal activities) to predict the criminal rank of a mobster.

While the studies above provided insight into the social organisation of and possible countermeasures against criminal organisations, the application of SNA to criminal groups nearly inevitably faces problems of *noisy or incomplete information*. Information on a criminal network is often likely to be missing or hidden, due to the covert and stealthy nature of criminal actions [97, 84]. Consequently, the derived networks are incomplete, incorrect, and inconsistent, either due to deliberate deception on the part of criminals, or to limited resources or unintentional errors by LEAs [98, 99, 100, 101, 60]. These limitations may bias the analysis and cause problems of uncertain information, potentially jeopardising the effectiveness of the investigations [85].

On the other hand, when seeking optimal disruptive strategies for criminal networks, two main approaches can be considered [102]: the *human capital* and the *social capital*. The former originates from economics, and refers to the personal attributes and/or resources possessed by actors within a social network. Sparrow [41] suggested that identifying the individuals who possess many resources and skills offers a great opportunity to damage the criminal network. Cornish [103] introduced the notion of *script*, which is borrowed from cognitive science. A script approach is a way to better understand how crimes are committed and how to prevent them. The central element of this approach, the crime script, is a step-by-step account of the actions and decisions involved in a crime. If the script is correctly identified, it can be used to prevent or disrupt crime commission. Later on, Bruinsma and Bernasco [104] combined this script concept with SNA, to identify the role of human capital within criminal networks.

By contrast, the social capital network-disruption strategy [105, 79] refers to the connections or ties between the actors in a network. It is through these connections that actors can have strategic positions, exchanging and sharing resources with

other actors in the network [41, 42, 93, 106, 102]. Research in this field is often based on SNA to find the most influential or powerful individuals of social capital, who correspond to the most central nodes in a network [107]. There is empirical evidence that *brokers* (*i.e.*, the individuals acting as bridges between disconnected subgroups) have a key role in the connectivity of criminal networks, often relating separate criminal collectives within illegal markets [93, 1, 108, 109, 99, 100, 101, 60]. For instance, the impact of brokers on the crime commission processes has been investigated by Morselli and Roy [108].

Through SNA, a number of interesting findings have been made over the years. Agreste *et al.* [60] devised an efficient approach for dismantling mafia syndicates, based on the application of percolation theory. Peterson [2] argued that the most central actors in covert networks might also be the most visible, and for this reason the most likely to be detected. Spapens [110] identified a brokerage role within Dutch ecstasy production, observing that brokers not only increase "social capital" within these criminal collectives, but also add "human capital". Bright *et al.* [111] investigated the effectiveness of five law enforcement interventions in disrupting and dismantling criminal networks, using both the social, and human capital approaches. Moreover, they showed how the removal of actors based on the Betweenness centrality metric was the most efficient strategy.

2.3.3 Sicilian Criminal Networks

As asserted in the previous sections, in recent years there has been a growing interest in the application of methods from Statistical Physics and Social Network Analysis (SNA) to the study of different kinds of crimes and, particularly, terrorism. We focus on a specific criminal organisation, the Sicilian Mafia (also known as *Cosa Nostra*), which originated in Sicily and has now spread worldwide [112, 92, 96]. Due to its global spread, Mafia controls entire economic sectors, influencing the social and political life of a country (*e.g.* by interfering in the results of electoral competitions).

In fact, Mafia tends to create deep roots into the very fabric of society, to the point that it becomes "impossible to destroy without a radical change in social institutions" (in the words of Italian politician Leopoldo Franchetti, 1876 [112]).

An impressive scientific interest for the study of social structure of Sicilian Mafia syndicates has been generated because of the social embeddedness of this type of criminal organisation [113, 114]. Each of these groups can be referred to as *cosca* (*i.e.*, a Sicilian word which refers to any plant whose spiny closely folded leaves

symbolise the tightness of relationships between members of the Mafia), *gang*, *clan* or *family*.

There is a vast number of studies of criminal organisations and, terrorist acts; yet Mafia has characteristics that make it unique. If we analyse terrorism, it is possible to see how the organisations are formed by individuals who collaborate to pursue an objective and are even willing to sacrifice their lives to achieve that goal (*e.g.* the Twin Towers attack). After reaching their goal, the terrorist organisation generally dissolves (*e.g.* the IRA in Ireland).

The *modus operandi* of Mafia is different. Indeed, Sicilian Mafia has a particular structure that differs from common criminal networks (such as the terrorist nets). The affiliates are bound by blind loyalty and they still pursue further goals even after achieving a previous one. Moreover, Families last for several generations. They also tend to diversify their objectives: from controlling entire economic sectors (*e.g.*, by giving "protection" to small traders and taking control of larger factories), to influencing countries political life (*e.g.*, by interfering in the results of electoral competitions). The blind loyalty of affiliates makes it even more difficult to obtain reliable information about those criminal networks topologies: important information about such criminal network is likely to be missing or hidden, due to the covert and stealthy nature of criminal actions [97, 84, 98, 99].

In addition, a Mafia clan lasts for several generations [115, 96], and is characterised by two key elements: the close links among affiliates and the ability to lead illegal activities to pursue specific objectives. In the first element, the links between mobsters are very close and are marked by reciprocal altruism. In the second element, the clan identifies objectives that it considers to be profitable and almost safe (*e.g.*, human trafficking), and focuses its resources on those objectives. As the objectives change over time (*e.g.* the members of the clan risk overexposure and capture, or if the deal is no longer profitable), the organisation defines new objectives and redesigns its structure to achieve them.

In complex networks terms, this is a rather unusual behaviour. which is almost never detected in other criminal networks. For this reason, many studies have been done on the structure and evolution of a Mafia syndicate (*i.e.*, a "cosca"). Yet, existing studies are mainly qualitative since very few (and restricted) datasets exist. We overcome this limitation by creating datasets that allow network analysis.

It has been challenging to define a new dataset directly derived from judicial documents. The information in these documents were verified by the police and

the magistrates during a trial, making this a reliable dataset. An additional challenge is that such datasets are bound to be incomplete, for instance due to gaps in the investigation process. A common case is when the police is not authorised to intercept a specific group of individuals during a certain period of time.

2.3.4 Resilient properties of Networks

Several authors [116, 117] have described the concept of network resilience considering two main aspects: (i) the capacity to absorb and thus resist disruption, (ii) the capacity to modify the network's internal structure and strategies, in order to adapt to external pressures. Resilience depends on the level of *redundancy* present in the criminal network. Redundancy is reflected in the number of relationships among the network actors, and is associated with strong connections between these actors [118]. A high level of redundancy and consequently of diversity of relationships in the network allows it to function even if some tie is broken and to find more options to substitute the network actors who have been arrested, incarcerated, or killed by LEAs. Replacements are often found within short social distances because criminal connections often start from already established social networks of kinship, friendship, or emotional ties [119, 113].

Network resilience can be analysed by node removal processes, which are also used to manage cascading failures (for instance in power grids) [120], confine the spreading of viruses [6, 7, 8, 9, 10, 11] or fake news [12, 13, 14], or streamline the network, through pruning, *i.e.* by removing those nodes that have minor impact on the overall connectivity [121].

A common strategy used to detect which nodes should be removed first is by ranking them according to centrality metrics. Typical metrics are the *degree*[30], the *h-index* [45, 122], the *coreness centrality* [123, 124], the *Eigenvector centrality* [48], and the *Katz centrality* [49]. Indeed, those metrics are a widespread tool in Network Science to identify important elements (*nodes* and *edges*) in real-life networked systems.[1] Yet, the efficiency of each metric (at ranking the network elements) tends to vary with the type of application. Thus, the choice of the metric underlying the node-removal process is often domain-specific and requires tailored experimentation.

[1]Other popular centrality metrics are *Betweenness centrality* and *Closeness centrality*; yet their computation requires to calculate all-pairs shortest paths in a graph, which is prohibitive even in graphs of modest size. Thus, in the work presented in Chapt. 5 Betweenness and Closeness centrality have not been considered.

2.4 Summary

In this chapter, we have provided a concise background of the main theoretical concepts that this thesis develops beyond the state-of-art. In particular, in Sect. 2.1 the Graph Theory main mathematical concepts were described: starting from the basic notions, moving forward to the definition of the artificial models, and concluding with the explanation of the most significant metrics herein used for conducting our experiments both in terms of computation and evaluation. Next, in Sect. 2.2 we moved our attention from the theoretical concepts of Machine Learning to the state-of-art in terms of how the academic community approached the Sparse Artificial Neural Networks field so far. Lastly, in Sect. 2.3 the literature of Network Science referred to such application domains (*i.e.*, Criminal and Artificial Neural networks) has been shown.

Chapter 3

Criminal Networks Part I: Disruption Analysis

The work included in this chapter has been published in the following papers: [16, 17].

3.1 Introduction

In this chapter, we borrow methods and tools from Social Network Analysis (SNA) to (i) unveil the structure and organisation of Sicilian Mafia gangs, based on two real Sicilian Mafia networks, and (ii) gain insights as to how to reduce the Largest Connected Component's size (LCC) of them. We focus particularly on one aspect of *network resilience* that, in this context, represents the ability of criminal networks to survive to (or counter) the actions of the Law Enforcement Agencies LEAs. Network resilience also relates to the capacity of these networks to reorganise after perturbations (*e.g.*, police raids), and to re-establish connectivity. However, this particular aspect has not been investigated herein because our networks are currently treated as a static dataset. Another view of network resilience will be discussed in Chapt. 5.

In Chapt. 2 we have been extensively mentioned that the Sicilian Mafia differs from other criminal organisations. Indeed, Mafia networks have peculiar features, due to the links distribution and strength, which makes them extremely robust to exogenous perturbations. Analysts are also faced with the difficulty 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. This aspect will be more deeply discussed in Chapt. 4.

A key feature of our work is the generation of two real–world datasets, which we have anonymised (to eliminate sensitive data) and made publicly available online on GitHub repository[1] and on Zenodo [125]. These are based on raw data derived from juridical acts, relating to a Mafia gang that operated in Sicily (Italy) during the first decade of the years 2000s. We created two very different networks, capturing phone calls and physical meetings, respectively that we have characterised in our previous conference paper [16]. Our datasets relate to a Mafia syndicate acting as a link between prominent criminal families, operating in the two biggest cities (Palermo and Catania) of Southern Italy. The Phone Calls (PC) dataset has been derived from eavesdropping, while the Meetings dataset (M) has been derived from police surveillance data. Both datasets are represented by undirected networks. For each of them we have created a weighted graph version (considering the frequency of interactions between individuals, or nodes), as well as an unweighted version (accounting only for connections). A detailed description of the datasets collection and management is given in Sect. 3.3.1.

The present study goes well beyond the initial characterisation of these datasets that we conducted in [16], which advanced the state-of-the-art through a creation of novel datasets from real–world data (two weighted undirected graphs), and an analysis that shows how the peculiarities (and internal dynamics) of Mafia families connections may be unveiled via SNA methods. Indeed, this work investigates network robustness across different scenarios, pinpointing the most effective metric, and demonstrating an effective strategy to obtain a faster LCC size drop. In detail, we simulate two types of police operations: (i) arresting one criminal at a time (sequential node removal), and (ii) police raids (node block removal). We evaluate how the different types of networks are impacted by these two types of perturbations, in terms of LCC size drop.

As asserted in Chapt. 2, the resilience of a social network is the result of several factors deriving from the network structure, such as the position of the nodes (*i.e.*, the social capital), and individual technical abilities (*i.e.*, the human capital): the former refers to the most connected actors (*i.e.*, key players) of a criminal network and/or the nodes connecting the several groups and subgroups inside a network (*i.e.*, bridges); the latter refers to the precious knowledge, skills and technical

[1] https://github.com/lcucav/criminal-nets/tree/master/disruption

abilities of a criminal network node (*e.g.*, pharmacological and chemical knowledge are required in drug synthesis processes).

We employ SNA methods to identify the actors having a high level of social capital. These are typically the most influential individuals, with a *central* role in the criminal network. To this end, we put to test four different centrality metrics, namely: (i) *Degree centrality*, (ii) *Betweenness centrality*, (iii) *Katz centrality*, and (iv) *Collective Influence*. It is worth recalling from SNA that the Degree centrality helps identifying network hubs (*i.e.*, the focal points). However, contrary to other types of networks, criminal networks communications are not necessarily mediated via hubs, which are more visible and, thus, more vulnerable [1, 2]. On the other hand, by weighing the communication paths (rather than nodes in isolation), Betweenness centrality pinpoints those nodes that play an important role in multiple communication paths. We have therefore hypothesised that Betweenness centrality could help removing the individuals that are crucial in maintaining the information network. In turn, removing those individuals would increase the LCC size drop inside the networks, which is our aim.

For the sake of completeness, we also considered two more prominent centrality metrics. Katz centrality computes the relative influence of a node, measuring the number of node's immediate neighbours (first degree nodes) and also all the other nodes in the network that connect to the node itself through these immediate neighbours. Finally, Collective Influence establishes the centrality of a node in a criminal network taking into account the degree of the node's neighbours at a given distance l from it.

Our analysis identifies Betweenness centrality as the most effective metric, showing how, by neutralising only 5% of the affiliates, the LCC size dropped by 70%. We also identified that, due the peculiar type of interactions in criminal networks (namely, the distribution of the interactions frequency) no significant differences exist between weighted and unweighted network analysis. Our work has significant practical applications for tackling criminal and terrorist networks.

3.2 Related Works

Villani *et al.* [126] tried to check whether, and to which extent, strategies based on both human and social capital could reduce (or neutralise) resistance and adaptation abilities of criminal organisations. Thus, we base our attack strategy on the social capital approach.

Inspired by the work of [126], we have also analysed the correlation between the social capital and the human capital in the criminal organisation we are considering (See Sect. 3.5.2).
In Sociology, the social capital denotes a group of tangible and intangible resources such as interpersonal relationships, a shared sense of identity, shared norms and values, trust, cooperation, and reciprocity that are fundamental to assure the functioning of human communities. From our experiments, as will be described in Sect. 3.5, individuals with high Betweenness contribute significantly to the social capital associated with a criminal organisation because they connect the various subgroups composing the criminal organisation itself and enable efficient information flow within the organisation.
Thereby, our experiments confirmed that the removal of high Betweenness individuals induces a strong reduction in the social capital but is not, however, the greatest loss that a criminal network might suffer from. For instance, in a drug-trafficking organisation, individuals with specialised knowledge in chemistry play a pivotal role to carry out the illicit affairs of the organisation. Thus, their removal seriously affects the organisation even if those individuals had low Betweenness.
Broadly speaking, the human capital refers to the skills and resources that some components of the organisations have and which play a key role in the functioning and long-term sustainability of the organisation. As noted in [126], the most effective repressive actions should simultaneously aim at weakening both the social and human capital of a criminal organisation. To this purpose, Villani *et al.* [126] suggested a new index (called CNR - *Criminal Network Resilience*) to assess the resilience of criminal networks, which combines parameters related to social capital, as well as parameters associated with the human capital.
In our study, individuals have three different roles, namely: leaders, members, and elements close to the gang but not affiliated with it.

3.3 Materials and Methods

This section describes the real criminal datasets (Sect. 3.3.1) we extracted from juridical acts. We used these datasets for our analysis, followed by a brief summary of the disruption strategy (Sect. 3.3.2). In particular, the experiments relate to a preliminary analysis on the networks typologies by the use of the weight distribution and the shortest path length that will be extensively discussed in Sect. 3.4. Then, once the network topology is well known, the disruption strategy could follow. The results obtained are separately discussed in Sect. 3.5.

3.3.1 Datasets Creation

In this section we explain the structure of the two networks taken into account (*i.e.*, Meeting and Phone Calls). This relates to our work published in [17], whereas the datasets are available on Zenodo [125]. Note that the work described in this section was conducted by the co-authors. Nonetheless, those information have been included to allow the reader to have a clearer overview on the whole process of this piece of research. Specifically, the author of this thesis dealt with the experimental part of this work in terms of implementation of the disruption strategy and interpretation of its results.

Dataset Collection: Our datasets were derived from the pre-trial detention order, issued by the Court of Messina's preliminary investigation judge on March 14, 2007, which was towards the end of the major anti-mafia operation referred to as the "Montagna Operation".
We have chosen the court order associated with the operation Montagna (a document composed by approximately more than 250 pages) because it contained the largest number of wiretaps and stakeout instances among all court orders we had access to, which allowed us to maximise the size of the dataset.
This operation was concluded in 2007 by the Public Prosecutor's Office of Messina (Sicily) and was conducted by the Special Operations Unit of the Italian Police (*Reparto Operativo Speciale* (R.O.S.) of the *Carabinieri*, specialised in anti-Mafia investigations).
This particular investigation was a prominent operation focused on two Mafia clans, known as the "*Mistretta*" family and the "*Batanesi*" clan. From 2003 to 2007, these families were found to had infiltrated several economic activities including major infrastructure works, through a cartel of entrepreneurs close to the Sicilian Mafia.
According to the Italian Code of Criminal Procedure, the pre-trial detention order begins with the crimes alleged against individuals. The same individual can be mentioned multiple times in the order: for example, an individual A could appear for the first time because she/he is accused of theft; furthermore, A could appear together with an individual B because both A and B could be jointly accused of extortion.
Information available in the pre-trial detention order is relevant to the construction of a graph, which falls within the class of the so-called of *co-offending networks* [127]: in a co-offending network, nodes are associated with individuals while edges specify that two individuals (corresponding to the endpoints of that edge) co-participated in perpetrating a crime. The construction of a co-offending net-

work highly depends on the assessments of the public prosecutor who includes in the court order only the activities having criminal relevance.
Each meeting has been represented as a clique with order equal to the number of participants. It may be a limitation to assume that each participant has spoken to all the others, but no more precise information can be gathered from the judicial proceedings. The phone calls networks is less dense than the meeting one because, even though the overall number of people involved is comparable, the calls took place between two individuals only. Furthermore, as highlighted in Table 3.1, there are only 47 people in common between the two networks and, thus, for a more complete analysis it is necessary to work with both the networks.

The pre-trial detention order reported the composition of both the Mistretta family and the Batanesi clan. Judicial proceedings followed a chronological narrative scheme, which detailed the illicit activities pursued by the members of the Mistretta clan before the imprisonment of a boss. To preserve anonymity, we will denote such a boss as x. The boss x has been selected by the most influential Mafia syndicates to settle the conflicts between the Mistretta and the Batanesi families. The conflicts were unleashed from the extortion imposed on local entrepreneurs in the construction of the highway that connects the city of Messina to Palermo. LEAs relied on the depositions of collaborators of justice, stakeouts and wiretapping. Meetings involving suspected individuals can be broadly classified as follows: (a) meetings which aim to define the organisational structure of the Mafia syndicate along with its relationships with entrepreneurs and other Mafia syndicates operating in surrounding areas. The boss x always attended these meetings and has always been accompanied by at least two trusted men. (b) Meetings involving people who occupied the lowest levels of the Mafia syndicate hierarchy; the purpose of these meetings was, in general, to design illicit activities and, usually, only two people were involved in these meetings. For each one the date of the meeting, the place, and the participants were recorded. Participants to a meeting were identified by a unique identification code.

The datasets were constructed by hand, transcribing the names of the subjects spotted in the meetings and intercepted in the phone calls. Of course, there are tools for the automatic extraction of information, but they operate satisfactorily only on structured documents (*e.g.*, in documents containing tables or records). However, the juridical acts under scrutiny is an unstructured document. Phone calls records are structured information; unfortunately, those kind of tools and records are not public and, thus, the appropriate software was only used by investigators. In addition, the information about the eavesdropping or stakeouts

Table 3.1: **Characteristics of Meetings and Phone Calls networks.**

Parameter	**Meetings**	**Phone Calls**
No. Nodes	101	100
No. Edges	256	124
Max. Weight	10	8
Max. Frequency	200	100
Avg. Degree	5.07	2.48
Diameter (Max. Shortest Path)	7	14
APL	3.308	3.378
CC	0.656	0.105
Common nodes	47	

are part of the arguments put forward by the prosecutor to prove, or attempt to prove, that a specific subject was associated with others to pursue a specific criminal goal. In the judicial act we analysed, only the significant individuals for the current investigation were included. Both eavesdropping and stakeouts must be authorised by the Judicial Authority for the purposes of the investigation and for a specific period of time. There are laws according to which, for example, interceptions that are irrelevant to the investigation must be destroyed and must not appear in court documents. It is also important to highlight that the judicial proceedings, from which we extracted the relevant data to build the two graphs, is a public document that any private citizen can request to read for consultation from the court, despite it contains names and telephone numbers (that can be considered as sensitive data). Then, the anonymisation process was performed for the sake simplicity as it is not easy to deal with nodes having so many details (like first and last names) as attributes and also for the sake of discretion. Hence, we replaced the names with progressive numbers (in order of appearance in the document) and, thus, two list of edges were obtained (one for each network).

The procedure to build the Meetings network was as follows: (a) each person who participated in at least one meeting corresponds to a node in the network; (b) two subjects in the Meetings network are connected by an edge if both of them attended at least one meeting; (c) edge weights reflect the number of meetings that two individuals jointly attended (*i.e.*, interaction frequency). Analogously, judicial proceedings recorded phone calls between the individuals under investigation. For each call, the ID of the caller (resp., the ID of the receiver), the time of the call, the duration, and content of the conversation were reported. The procedure to build the Phone Calls network was as follows: (a) each person who made or received at least one call was associated with a node in the network; (b) two nodes in the phone call are connected by an edge if the associated individuals had at

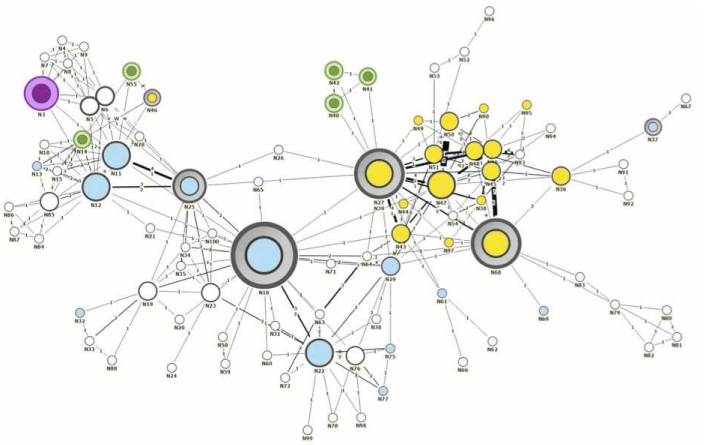
least one call; (c) the weight attached to an edge specifies the number of phone calls between the individuals connected by that edge. Observe that we used only text data (from pre-trial detention order) to derive links in both the Meetings and the Phone Calls networks. In addition, we included all actors found in the investigation records, independently of the fact that an actor is actually a member of a Mafia syndicate or not.

Dataset Description: Figure 3.1 is a visual representation of both datasets as undirected and weighted graphs. In our datasets, each node represents an individual, whereas links show mutual interactions among the individuals. The Meetings dataset, captures the physical meetings among suspects. The Phone Calls dataset refers to phone calls among individuals. As previously asserted, the datasets are available on Zenodo [125] and were discussed in detail in our earlier studies [16, 17, 81].

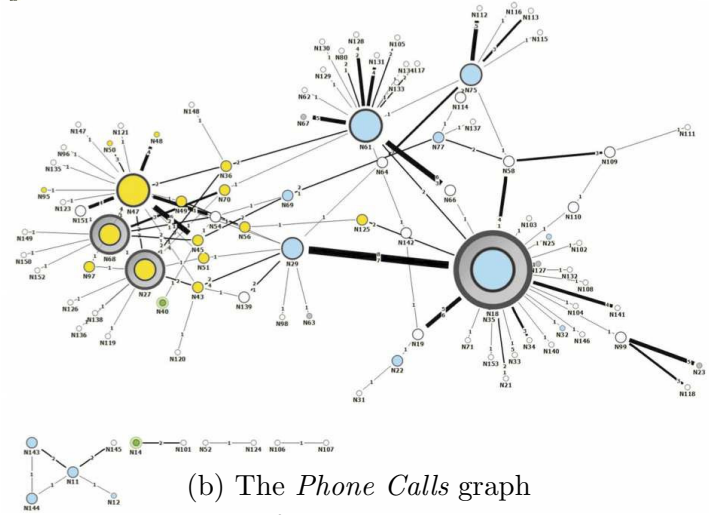
The key features of our datasets are summarised in Table 3.1, including the number of nodes and edges, the maximum weight, the maximum frequency in the affiliates' interactions, the average degree (neighbours per node), and the highest number of individuals required to connect mobsters (based on all shortest paths in both datasets). We also calculated the APL and the CC of the Meetings and the Phone Calls networks (see Sect. 3.4.3). It is worth noticing that the two datasets have 47 nodes in common and a comment on this overlap is extensively discussed in Sect. 3.6.3.

Indeed, both networks can be treated as either unweighted and weighted networks since, for each pair of individuals, we recorded a coefficient representing the number of times the pair had a meeting (as reported by the police surveillance logs), and made a phone call (reported in police interceptions logs). In SNA terms, these coefficients are known as the strength of the connection between two individuals.

Nodes may belong to different categories. Some nodes represent the leaders (*i.e.*, "bosses") or the soldiers (*i.e.*, "picciotti", a Sicilian word that refers to the lowest rank of the Mafia hierarchy) of the criminal organisation. Other nodes are associated with individuals (*e.g.*, a fruit seller or a baker) who had one or more calls with members of a Mafia syndicate but are not affiliated with any criminal organisation.



(a) The *Meetings* graph



(b) The *Phone Calls* graph

Figure 3.1: The graphs derived from the juridical acts data extraction. The colours represent the different clans. In particular, turquoise nodes represent the members of the "Mistretta" family, while the "Batanesi" family is drawn with yellow nodes. Circled nodes correspond to leaders (*i.e.*, bosses) investigated for having promoted, organised, and directed the Mafia association. The green and purple circled nodes refer to bosses of Mafia families of other mandates. Finally, the white nodes represent other subjects who are close to a family, but are not classifiable in any of the previous categories. In both graphs, the edges width is proportional to the number of meetings or phone calls, and the size of the nodes is proportional to their degree.

3.3.2 Disruption Strategy

This section briefly describes the experimental process employed to disrupt the two networks and evaluate the effects of node removal, under different conditions and strategies.

Design of the Experiments: Both datasets were used, under both unweighted and weighted conditions. Two node removal strategies have been studied (*i.e.*,

sequential and block). These are iterative procedures in which the nodes have been removed in decreasing order of their centrality score. After the node removal stage, the LCC size is updated and the process resumes.

Let us denote by $LCC(G)$ the size of the LCC of a graph G. Denote by G_i the graph resulting after the i-th iteration of the node removal algorithm, whose size of the LCC is $LCC(G_i)$. Clearly, the unperturbed graph is G_0, for which we have $LCC(G_0)$.

The relative difference between the size of the LCC at the start of the simulation and after the i-th iteration (*i.e.*, i-th node removal) is given by $\rho_i \in [0, 1]$:

$$\rho_i = 1 - \Big|\frac{LCC(G_i) - LCC(G_0)}{LCC(G_0)}\Big| \tag{3.1}$$

Note that $\rho_0 = 1$ and $\rho_n = 0$, where n is the last iteration (sequential removal).

Both strategies (sequential and block removal) may be summarised as follows:

1. We first compute $LCC(G_0)$; *i.e.*, the LCC size for the initial graph G_0.
2. This step depends on the removal strategy. Either the the highest ranking of the remaining nodes (in the sequential strategy), or the the set of the five most influential nodes of the remaining ones (in the block strategy) are removed. Ranks are computed with the current centrality score. The new graph G_i, with $i = 1, 2, \ldots, n$ is obtained (for block removal we have fewer iterations).
3. Compute $LCC(G_i)$, and calculate ρ_i.
4. Steps 2 and 3 are repeated until the graph size can no longer be reduced.

Sequential Nodes Removal: It simulates the scenario in which affiliates are arrested one-by-one by the police.

Block Nodes Removal: It simulates the scenario in which affiliates are arrested during a raid by the police. This strategy is similar to the sequential one, with the main difference being that nodes are removed in blocks of five. The block size depends on the type and scale of the datasets. The fraction of nodes to be removed during block police operations is a reasonable value that takes into account some considerations. In [60] a serious reduction of the LCC with only 5% as block size was obtained. Moreover, in such a relatively small criminal network

larger fractions of block sizes appear unrealistic. This is why, in our case, five has been found to be adequate in terms of number of nodes removed at once.

3.4 Networks Characterisation

This section describes our work published in [16] that relates the weight distribution (Sect. 3.4.1) and the shortest path length (Sect. 3.4.2) analyses.

3.4.1 Weight Distribution Analysis

We begin this study by discussing the edges weights distribution in both *Meetings* and *Phone Calls* networks. Fig. 3.2 shows the weight distribution that specifies, respectively, the amount of meeting and phone calls exchanged between pairs of individuals in the networks. On the horizontal axis we report edge weights, while the vertical axis shows frequencies.

Noticeably, both networks exhibit similar characteristics and include several low-weight links. Thus, there are just a few high-weight edges; *i.e*, nodes incident on those links exhibit an high number of interaction within the network. A possible explanation is that the affiliates want to reduce the risk of being intercepted by law enforcement, and even by other people outside the clan. In the *Meetings* network this trend is even more accentuated (the maximum frequency in low-weight links is almost double, as shown in Table 3.1). Moreover, the maximum weight of interactions among affiliates in the *Meetings* network (*i.e.*, $w = 10$) is greater than the one in the *Phone Calls* network (*i.e.*, $w = 8$). A possible explanation is that mobsters prefer to communicate by physical meeting rather than calling each other, to reduce the risk of being intercepted by the police. Mobsters will find it easier to crypt their conversations in face-by-face meeting, for instance by using body language, or generating background noise. Furthermore, bosses often have to participate to Mafia events to pursue their power inside a clan. For instance, bosses have to participate to funerals of other affiliates, and other solemn religious demonstrations (masses, processions, etc.). During those kinds of events, they also have the opportunity to pass messages to their closest subordinate affiliates. Moreover, it is harder for criminals to notice that they are going to be intercepted rather than to be eavesdropped by the police.

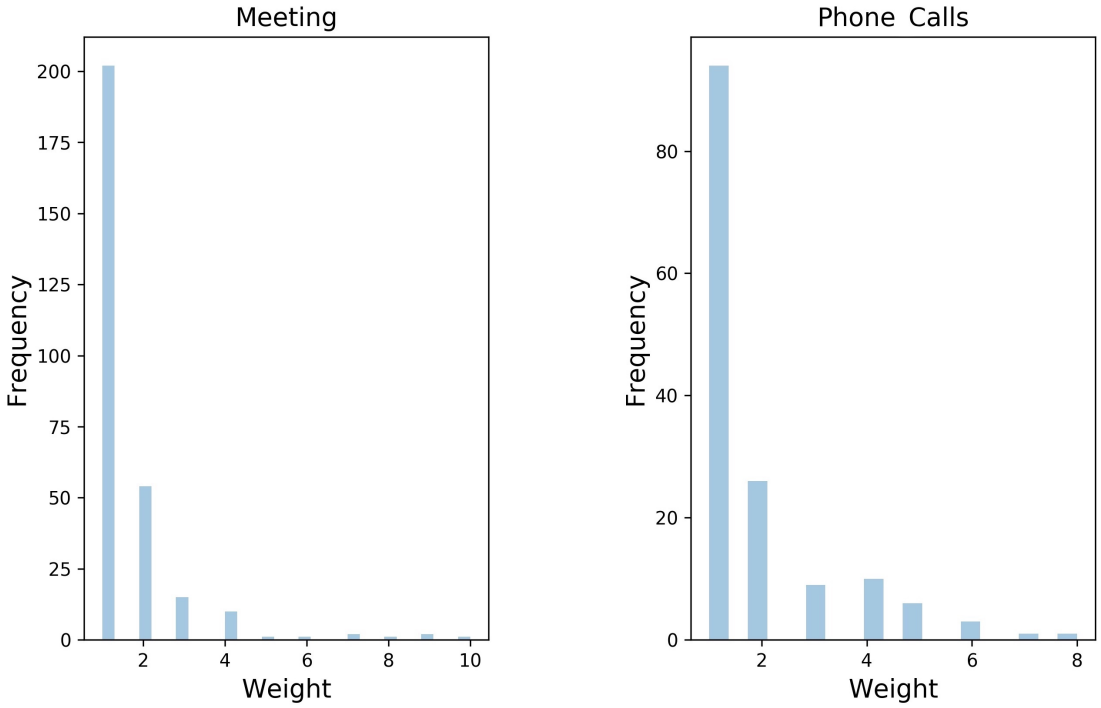


Figure 3.2: Weights distribution. **Left Panel** *Meetings* network. **Right Panel** *Phone Calls* network.

3.4.2 Shortest Path Length Analysis

The histograms of shortest path length distributions of Fig. 3.3 provide useful statistical characterisations of the two networks under scrutiny. Path length statistics are closely related to dynamic properties such as velocities of network spreading processes. Usually, criminal organisations are structured in a way as to optimise the number of communications among members, and to efficiently disseminate information. These members can be discovered by following short paths of communications. Moreover, we can discover relationships among individuals belonging to distant groups in the graph because, even when two nodes seem to be distant, there may exist a relatively short path that connects them; *i.e.*, affiliates may also be acting as a *bridge* to connect distant groups in the network.

There are similarities between the weighted and the unweighted shortest path length analysis. In both scenarios, indeed, there is a higher interaction frequency among affiliates having a balanced number of intermediates. This means that they do not like to spread their encrypted messages with a too low (resp., high) number of intermediates. This behaviour confirms the hypothesis that inside a "cosca" it is better to avoid the borderline cases. On one hand, if the shortest path is composed of a lower number of affiliates, the bosses are overexposed to police investigations. On the other hand, the longest the number of intermediates,

the higher the chances to be intercepted by people outside the Family.

Furthermore, in the weighted simulations emerges a lower frequency of interactions in the *Phone Calls* dataset compared with the same shortest path length of the *Meetings* network. This behaviour, emerged also in Fig. 3.2, proves that the clan tries to minimise the risk of interceptions, specially to avoid exposing those mobsters who are hierarchically in a higher rank.

The availability of a real weighted graph is a valuable asset in order to conduct a more thorough network analysis. Indeed, in the unweighted scenario this behaviour is not highlighted because both datasets seem to act in the same way.

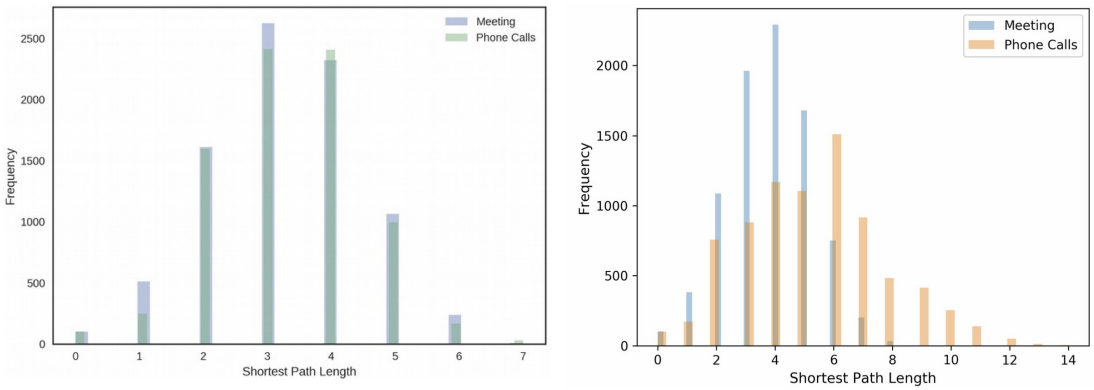


Figure 3.3: Distribution of shortest path lengths in *Meetings* e *Phone Calls* networks. **Left Panel** The *unweighted* graph. **Right Panel** The *weighted* graph.

3.4.3 Average Path Length and Cluster Coefficient Analysis

The APL and CC values in the Meetings network were compared with the average values of APL and CC of a random graph $G^{\star}$. In our tests, $G^{\star}$ had the same number of nodes and the same average degree of the Meetings network. In the Meetings network, the APL and CC are equal to 3.308 and 0.656, respectively; in the random graph associated with the Meetings network, we measured an APL and a CC equal to 4.342 and 0.0018, respectively.
We repeated the procedure above for the Phone Call graph and we generated a random graph with the same number of nodes of the same average degree of the Phone Call graph and we calculated its APL and CC. We observe that the APL and the CC for the Phone Calls network are equal to 3.378 and 0.105, respectively; in the random graph associated with the Phone Calls network, we measured an APL equal to 5.593 and a CC equal to 0.0039. We can therefore conclude that both in the Meetings and Phone Calls networks the CC values are considerably

bigger than those found in the corresponding random graph. In contrast, APL values are moderately lower than those found in the corresponding random graph.

In the light of our analysis, both the Meetings and the Phone Calls networks can be regarded as *small world networks* [39]. The high CC coupled with a low average path length favour the information flow among the individuals in the criminal organisation and promote their coordination, thus making the organisation more effective.

3.5 Disruption Results

In this section the results obtained from our experiments on LCC size are shown. The network analysis is reported next, considering weighted and unweighted graphs configurations (Sect. 3.5.1). Next, in Sect. 3.5.2 there is a quick discussion in comparing our analysis with the role of the criminals within the network (*i.e.*, human capital). Lastly, Sect. 3.5.3 will summarise the most important results obtained from our analysis.

3.5.1 Weighted Graphs Analysis

Fig. 3.4 shows the results obtained for the cases of sequential and block node removal, for both datasets, and including all four centrality metrics. Remarkably, the Katz coefficient (tuned to the default values of $\alpha = 0.1$ and $\beta = 1.0$) is the least effective one (i.e, the slowest one) at causing the faster LCC size drop, in all eight cases, that are: two datasets (Meetings and Phone Calls), two strategies (sequential and block) and two graph structures (weighted and unweighted).

To understand this result intuitively, we need to look at the way this centrality metric operates. Katz determines the importance of each node based on the number of *walks* that pass through it, but it does not consider their length. Furthermore, shortest paths are not considered, hence a walk may visit the same node multiple times. Yet, this is in contrast to how criminals would operate in practice. Affiliates typically prefer to spread the information through a number of intermediaries, to minimise the risk of interception by non-family members. This is consistent with our earlier findings [16] (See Sect. 3.4). Ultimately, it would not make sense (and would be unwise) to send the same message multiple times through the same path, which is what Katz would help identifying. Therefore, removing nodes by the highest Katz score would not be a winning strategy.

All the other metrics act better than Katz centrality, and comparably among each other. This happens because of the weights distribution shape [16], which exhibits a long tail of nodes, with just a few dominating ones (Sect. 3.3.1). Thus, after removing the most central nodes (*i.e.*, the first five iterations), the network gets almost totally disconnected and the remaining nodes have the same weight ($w = 1$). Hence, all the metrics focused on either degree (*i.e.*, Degree and Collective Influence) or shortest paths (*i.e.*, Betweenness) follow the same ρ drop speed. On the other hand, Katz centrality with its default parameters focuses on walks of undefined lengths, thus producing a slower ρ drop.

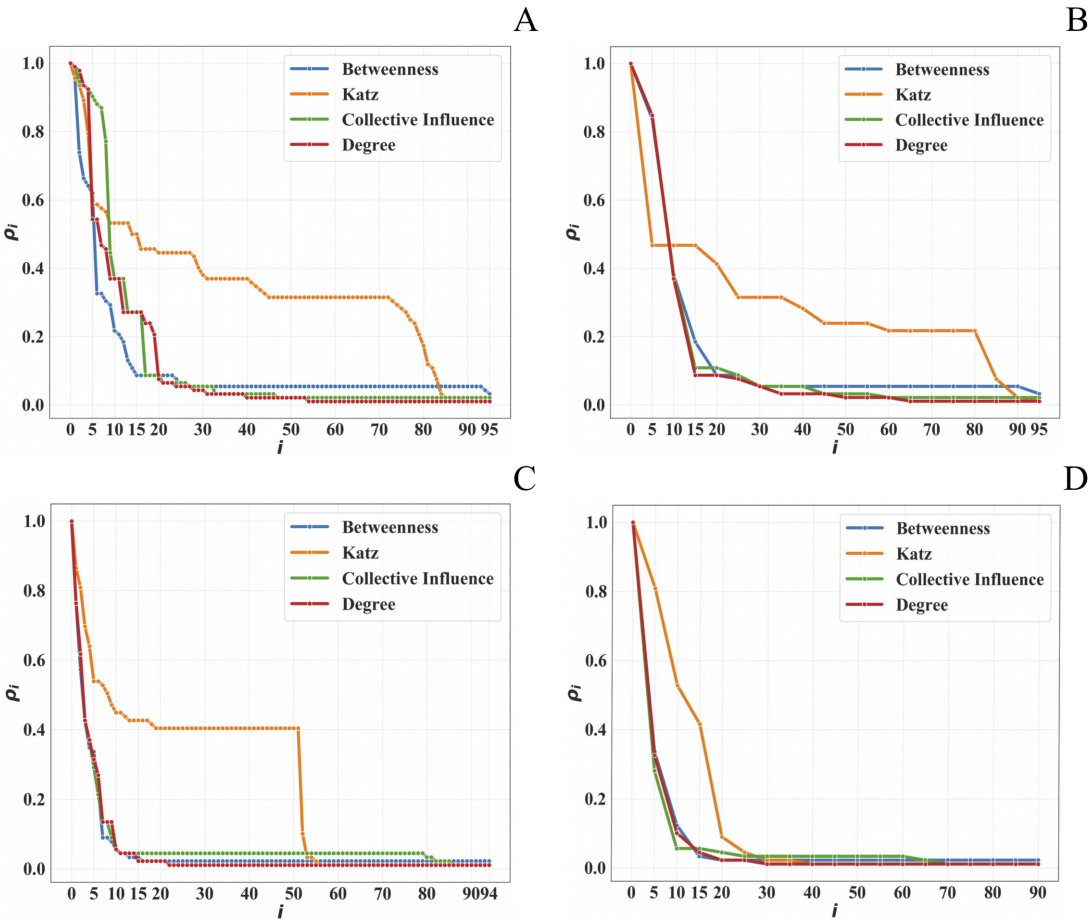


Figure 3.4: **Weighted networks.** A: Meetings dataset, sequential node removal strategy. B: Meetings dataset, block node removal strategy. C: Phone Calls dataset, sequential node removal strategy. D: Phone Calls dataset, block node removal strategy.

Sequential vs Block Removal: Looking at Fig. 3.4, with the exception of Katz, no significant differences are visible between the two node-removal strategies (*i.e.*, sequential and block). This is somewhat counter-intuitive, since in real life police raids are typically aimed at breaking up the network more effectively.

In our case, this result originates from the particular type of the datasets at hand. When constructing the datasets, we did not have access to information about the way criminals reconstructed their communication channels following arrests. Hence, our network is static (*i.e.*, it misses the network reconfiguration data), which is why our analysis is not fully capturing the dynamic aspects that differentiate sequential and block strategies. In network terms, this translates into no differences in terms of LCC size drop as the network is static. On the other hand, significant network re-tuning of nodes importance, due to internal re-organisation of trusted affiliates used to spread messages within and outside the criminal network, would be expected in the case of dynamic graphs (*i.e.*, graph snapshots before and after police operations).

Weighted vs Unweighted: Considering now the differences between weighted and unweighted graph analysis, we notice that the majority of cases do not pin-point major differences. This was due to the peculiar way in which weights are distributed in criminal networks (as noted in the *Weighted Graphs* paragraph).

Nevertheless, interesting differences are visible in the Meetings dataset - sequential node removal (Fig. 3.5). The unweighted case is mostly faster than (although occasionally equivalent to) the weighted case. This is because the weights (*i.e.*, the affiliates' interaction frequency) are concentrated in very few individuals, with most other weights having $w = 1$. This is also why, with the exception of the initial transient period (involving very few interactions), most algorithms converge to similar values.

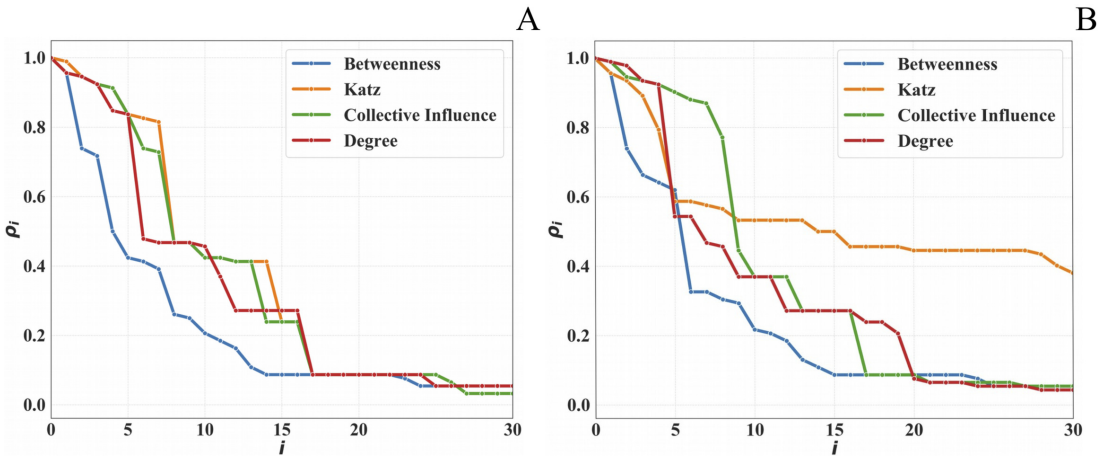


Figure 3.5: **First 30 iterations of the sequential node removal strategy, Meetings dataset.** A: Unweighted Graph. B: Weighted Graph.

3.5.2 Correlation between social capital and human capital

This section shows a comparative analysis of the top 10 nodes (*i.e.*, criminals) of the Meetings (resp. Phone Calls) graphs, sorted by decreasing values of the best centrality metric able to cause the highest fragmentation within the criminal network under scrutiny (*i.e.*, the Betweenness centrality), and the role of such nodes within the organisation. The roles we have considered are: leaders, members, and elements close to the gang but not affiliated with it. Leaders contribute more significantly than others to human capital because they are in charge of making decisions, planning and coordinating criminal actions. In turn, the human capital associated with the members of the gang is greater than those who are close but not affiliated with it.

From Table 3.2 (resp. Table 3.3) we observe that the node with ID 18 is associated with a leader and has the highest Betweenness, both in the Meetings and in the Phone Calls networks.
In addition, many nodes (18, 27, 29, 36, 47, 68) have high centrality, in both the Meetings and in the Phone Calls networks. We underline that both in the Meetings and in the Phone Calls networks there are only two leaders in the top 10 positions, which indicates that there is a weak correlation between social and human capital, and this agrees well with the results of [126].

Table 3.2: **Betweenness centrality values and Role (Meetings network).**

Position	Node ID	Betweenness centrality	Role
1	18	0.373	Leader
2	47	0.22	Member
3	27	0.159	Leader
4	68	0.126	Member
5	12	0.117	Member
6	25	0.114	Leader
7	29	0.09	Member
8	36	0.072	Member
9	22	0.069	Member
10	11	0.063	Member

3.5.3 Summary of the main results

This section provides a brief recap of the most important results and considerations.

Table 3.3: **Betweenness centrality values and Role (Phone Calls network).**

Position	Node ID	Betweenness centrality	Role
1	18	0.418	Leader
2	61	0.282	Member
3	47	0.236	Member
4	29	0.22	Member
5	75	0.096	Member
6	36	0.085	Member
7	27	0.083	Leader
8	68	0.068	Member
9	58	0.06	close
10	70	0.051	close

The best centrality metric: Comparing the algorithms in Fig. 3.5, it emerges that Betweenness centrality is by far the most effective centrality index for reducing the size of the LCC of a criminal network.
This result is consistent with literature reports upon criminal networks' SNA shown in the Introduction, and has an intuitive explanation. Indeed, to avoid being intercepted, members of a criminal networks build their relationships to assure that information flows along the shortest possible paths. In this way, both the Meetings and the Phone Calls network configure themselves as small-world networks with a low average path length and a large CC. The nodes that intercept most of these shortest paths are those having the largest values of Betweenness centrality, and act as intermediaries to assure the quick flow of information from any source to any target in the graph.

To confirm this intuition, we progressively removed nodes according to their Betweenness centrality and we measured the corresponding variation of APL and the number n_c of connected components (see Fig. 3.6). These plots indicate that the selected removal of nodes amplifies the average distance between any pair of nodes in the Meetings/Phone Calls networks and, simultaneously, it creates an increasing number of disjoint components. A repressive action aimed at removing high Betweenness nodes has, therefore, a devastating impact on network topology because it causes an LCC size drop, as we observed a fast drop in ρ. Also, since the Katz centrality prioritises those nodes crossed by a large number of walks of arbitrary length, it is less effective in detecting the nodes acting as intermediaries, and whose removal reduces the LCC size the most.

Intuitively, Betweenness centrality outperforms the other metrics, thanks to its operation on paths, rather than on individual nodes degree. This is particularly

effective in criminal networks that are devised in such a way as to minimise the path length, in order to reduce the risk of police interceptions. Betweenness centrality compromises the most influential paths, leading to a faster drop in ρ. This feature is what makes Betweenness somehow opposite to Katz centrality (whose goal is to explore walks).

Collective Influence was the second worst performer after Katz. This is, again, due to its emphasis on node degree instead of path length. Collective Influence is also showing some differences between the weighted and the unweighted processes, exhibiting a lower ρ drop in the weighed graph. A possible explanation is that the weighted case identifies as influential nodes not only those with higher weights on the incident links, but also the nodes having high-weight only on immediate neighbours. This could reflect a typical situation in criminal networks, whereby the top-leaders avoid direct exposure and mediate all communications through a single trusted individual (or very few of them). On the other hand, this particular aspect is not detectable in the unweighted analysis.

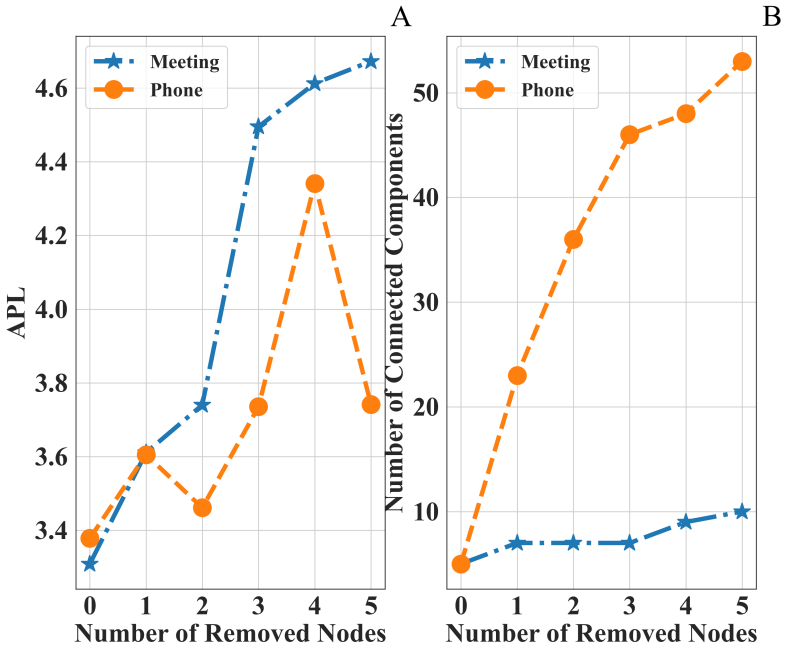


Figure 3.6: **Function variations of the number of removed nodes in the Meetings and Phone Calls networks. Nodes are prioritized on the basis of their Betweenness centrality.** A: Variation of the APL. B: Variation of the number of connected components.

Take-Home Message: In short, our results confirmed the effectiveness of SNA in speeding up the process of reducing the LCC size of criminal networks.

Considering our datasets, we could severely affect LCC (with a 70% LCC size drop) by neutralising less than 5% of the affiliates (either through sequential arrests or police raids). Betweenness centrality performed significantly better than the other three metrics, thanks to its specific focus on paths, rather than simple node degree. This is consistent with the typical operation of criminal networks where information diffuses through the shortest paths and within the organisation, to minimise intra-affiliates interactions and, thus, the risk of interception. Therefore, law enforcement interventions should favour path-related centrality metrics (such as Betweenness) instead of other strategies.

3.6 Discussion

In this section, the results obtained will be discussed with a view to the practical implication of the criminal networks under scrutiny. In details, Sect. 3.6.1 will relate the main outcomes obtained from the preliminary analysis we conducted in Sect. 3.4. Next, the effects of the disruption strategy will be commented in Sect. 3.6.2. Lastly, a brief summary of the main liming factors we encountered in this work will be presented in Sect. 3.6.3.

3.6.1 Characterisation Outcomes

The preliminary characterisation related the weight distributions and the shortest path length analysis was conducted. The weights distribution analysis unveiled a long tail of nodes, with just a few dominating ones. We could observe that a predominant number of links had similar weight ($w = 1$), thus making the distinction between weighted and unweighted graphs negligible after the removal of the best connected nodes. As also confirmed through the shortest path length analysis, the study demonstrated that just a few affiliates tend to be responsible for the highest interaction frequency with a balanced number of intermediates. This is because the clans aim to avoid an overexposure of the bosses (and other prominent affiliates), to minimise police interceptions.

Herein, we have moved our attention to also understand how LCC varies as a result of police interventions (*i.e.*, the nodes removal process).

3.6.2 Disruption Outcomes

Our social capital investigation shows that Betweenness centrality is the most relevant centrality metric, as it causes an effective LCC fragmentation for both networks under scrutiny. Indeed, Betweenness is the only metric (out of the four analysed) that focuses on shortest paths, which reflects the structure of Mafia syndicates. The LCC size drop velocity depends on the capability of centrality metrics to target the appropriate node with the right criterion, in terms of node importance, for the specific network topology. Mafia syndicates topologies are based on trusted affiliates to spread messages (*i.e.*, interact on shortest paths). Thus, it became clear that centrality metrics based on shortest paths led to faster LCC size drops.

Interfering on the paths produces a sensitive LCC size drop among trusted affiliates. The same conclusion has been drawn regardless of weights (to reflect interaction frequencies).

Overall, we could substantially and more rapidly reduce the LCC size by removing the top 5% most influential nodes, computed according to Betweenness score.

This effect is achieved thanks to the network weights distribution (or rather, their concentration within few influential nodes), which leads to a rapid drop in ρ. Once the most influential nodes have been removed, the remaining ones are largely characterised by $w = 1$, which makes the weighted and unweighted networks virtually indistinguishable.

Our SNA results can be directly translated onto law enforcement actions, considering that we are now able to efficiently identify the top 5% most trusted affiliates (*i.e.*, the ones typically employed as intermediaries between bosses and the other members).

In turn, we can virtually neutralise the clans' internal communication infrastructure by getting the trusted affiliates in custody. Intuitively, whenever arrests can be made in block (raids), that would further impair the ability of the criminal communication network to be re-established. However, we have not studied this specific aspect, due to unavailability of necessary data.

The pre-trial detention order is the final outcome of a time-consuming police investigation. Once the inquiry is underway (and even before it has been completed), the actual network is richer than the one derived from the pre-trial detention order. The investigative network includes extra interactions among suspects, which are removed once the judge deems them to be irrelevant. Thus, the final network

derived from the original one is partial and misses the data included in the initial investigation. This explains our limits related to the lack of data.

Furthermore, when LEAs inspect on those kinds of criminal networks, most of the time they have prior knowledge thanks to criminal records, even though they may not have a clear picture of the connections between individuals.

Generally speaking, the aim of a "cosca" is to conduct illegal activities (which may vary from place to place and are susceptible to local trends), and to ultimately pursue effective financial benefits. For instance, some clans may focus on drugs, rather than organ trafficking, prostitution, finance, or political influence. Quite commonly, clans pursue multiple activities, which makes it even more difficult to reconstruct the labyrinth of criminal communication networks (and perform SNA thereof).

Our datasets emerged directly from a collection of official juridical acts, and focused on a single criminal activity (the securement of public procurement contracts). This involved a network of entrepreneurs, was confined to a specific geographical area, and captured information over a limited time span. The peculiarity of our networks is that the gang was established in relation to a specific event (procurement of a methanisation process), so it was not a pre-existent organisation. Thus, our dataset captured a relatively simpler snapshot of the complex entanglement of mafia criminals, which constitutes both a strength, and weakness of our study.

Indeed, if LEAs have prior knowledge, then our approach is even more efficient; otherwise, as is the specific network herein considered, two main issues may arise to conduct investigations: (a) *noise*, and (b) different organisational *time scales*. By noise, we mean that LEAs could have too much information (*e.g.*, too many interceptions or surveillance logs, some of which are worthless). By different organisational time scales we indicate that criminals already know how to contact a specific criminal for their illicit purposes (*e.g.*, a sniper), in a way LEAs might not be able to identify. Thus, they have to spend more time to reconstruct the inner relationships by exploring the evidences and, as previously asserted, this is a time consuming process.

On the one hand, the scope of our SNA is limited by the significance and breadth of the datasets at hand. We have mentioned already how a more dynamic analysis of the network could not be done in this case, as for instance, understanding the reconnection ability following events like individual arrests or police raids. Also, we

are capturing a single criminal activity in a confined spatio-temporal context. So, it was not possible to detect a broader and more diversified set of communications, such as those taking place in a more complex, multi-activity network. Nor could we detect external communications, such as those involving people who were not directly members of the criminal nets except for entrepreneurs (*e.g.*, politicians, magistrates and businessmen).

On the other hand, the greater specificity of our networks allowed a cleaner analysis, focused on unveiling some hidden communication mechanisms. Having reduced the parameters under scrutiny and the complexity of the system, we could pinpoint a simple, yet effective strategy for unsettling the connectivity of the network through a dramatic drop in the LCC size. This might have not emerged from the analysis of a more complex network. Also, this simpler framework has allowed us to swiftly test out our hypothesis and to obtain reliable results.

This could have been a challenge on a complex network, especially when multiple criminal activities take place in parallel.

3.6.3 Main limiting factors of our work

We conclude this section by describing the main factors that posed some limits to our study.

1. *Wiretapping is allowed strictly under well-motivated circumstances and for short periods of time, thus limiting the ability of police forces to gather data on Mafia activities.* According to the Italian Code of Criminal Law, LEAs should request the Public Prosecutor an authorisation to wiretap conversations, also by means of electronic systems if such an activity is indispensable to prevent crimes. If authorised, LEAs are allowed to wiretap individuals for at most 40 days, which may be extended by up to 20 days. If the Public Prosecutor grants an extension to wiretapping (and, in general, to other surveillance activities), she/he has to produce a written report in which the reasons justifying the prolongation of surveillance activities must be clearly explained. Therefore, LEAs have a limited ability of collecting the data they need to reconstruct the topology of a criminal network and all information produced during non-authorised periods are lost.

2. *Collaborators of justice are often the main source of information for the investigators; yet they cannot always be deemed to be reliable.* Italian LEAs make an extensive use of the depositions of collaborators of justice, *i.e.*,

criminals who abandoned a Mafia syndicate and decided to cooperate with criminal justice authorities in order to reveal the composition and organisational structure of a Mafia syndicate. Collaborators of justice are widely considered as a powerful tool in dismantling Mafia syndicates, but their credibility must be carefully checked. In the Montagna operation, there was a collaborator of justice who provided very accurate and detailed information in the early stages of the investigation tasks. Nevertheless, from a certain point onward, that collaborator was no longer considered to be credible and the information he provided was deemed as unreliable.

3. *Mafia leaders avoid using phones to communicate.* Investigations show that the affiliates of a Mafia syndicate are highly suspicious of being tailed by police and, thus, they carefully avoid conversations on cell phones, whenever possible. This implies that the Phone Calls network yields a partial reconstruction of the information flow in a Mafia syndicate. A further consequence arising from the low propensity of some subjects to make/receive phone calls is the low degree of overlap between the Phone Calls network and the Meetings network. In practice, criminals prefer alternative means of communication (*e.g.*, they use intermediaries to convey encrypted messages). Consequently, the Meetings network includes individuals who had never been intercepted by the police forces. In contrast, some eavesdropped individuals had no ties with the Mafia syndicate but are acquainted with some Mafia syndicate members (for instance, because of work, or other family reasons) and they had conversations with them. Therefore, these individuals have never attended any meeting and are excluded from the Meetings network.

3.7 Conclusions

In this chapter, important improvements compared to the state-of-the-art have been discussed. The first challenge was the generation of two real–world datasets, capturing the interaction among Sicilian Mafia members relating to a "cosca" that operated in Sicily (Italy) during the first decade of the 2000s. Specifically, our two datasets were derived from original juridical acts about two Sicilian clans who sought illegal profits from public procurement proceedings that have been validated with law and law-enforcement experts. What makes this work unique is also a quantitative study on the unusual interactions among the clan affiliates. Indeed, a "cosca" network acts differently from other criminal organisations (*e.g.* terrorist affiliations). Individuals tend to periodically re-aggregate in pursuit of

changing goals and to survive over time. There are cases in which the same goal persists for several generations. Whereas in other occasions goals will change rapidly, depending on external socio/economic/legal changes.

To facilitate the reproduction and further extension of our work, we have placed an anonymised version of the datasets and the source code in the public repository[1] and with citable DOI on Zenodo [125], including the Meetings dataset (constructed from police stakeouts) and the Phone Calls dataset (derived from police wiretaps). We have also derived weighted and unweighted versions of the datasets.

The main focus was the social capital analysis, investigated by the drop in the size of the LCC of the networks. But, firstly, a network characterisation was essential to understand the criminal networks topology under scrutiny.

Thus, from the weight distribution of the two datasets we could figure out which individuals called or met more often. While each connected pair provides evidence of at least one interaction, the edge weights have proved invaluable in unveiling the most significant connections, both within and outside Mafia families.

The comparative analysis of shortest path length between weighted and unweighted graphs shows how Mafia members favour indirect communications through a well-trusted set of intermediaries. Also this form of communication effectively spreads information among key affiliates. Moreover, the weighted analysis gives us more accurate results than in the unweighted analysis. Indeed, the frequency in the *Phone Calls* dataset in the weighted scenario is lower than the *Meetings* one. This highlights how the clans succeed in reducing the risk of being intercepted by the police. This behaviour is masked in the unweighted analysis.

We have, then, explored mechanisms required for identifying key individuals in the network and, in turn, speed-up the LCC size drop through minimal node removal. We considered two strategies, namely: (i) a sequential node removal approach, and (ii) block removal. The first one simulates the scenario in which the police arrest one "cosca" affiliate at a time. The second one, mimics a police raid. Next, we put to test four different centrality metrics, namely: (i) Degree centrality, (ii) Betweenness centrality, (iii) Katz centrality, and (iv) Collective Influence. The effectiveness of the centrality metrics has been validated trough the ρ parameter, which measures the drop of the LCC size, after node removal, compared with the initial LCC size.

Our experiments unveiled Betweenness as the most effective metric. Indeed, it produced a greater impact in terms of LCC size drop rate, thanks to its priori-

tisation of communication paths, rather than by individual nodes degree. Thus, the resulting optimal strategy was to order nodes by Betweenness centrality score and to remove them by a decreasing order of these scores. This procedure tackled directly the communication mechanisms used by criminal networks, which are designed to minimise the probability of interception by the police.

3.8 Future Work

This work is prone to considerable extensions and adaptations. For one, it would be interesting to conduct a more in-depth comparative study between social and human capital in Mafia associations. In fact, the resilience of criminal networks also depends on the personal qualities and competences of their members. In Network Science, those competences are represented as node labels describing node roles. Therefore, in order to better assess the strength of a criminal organisation, we should also look at the human capital endowment.

While we have looked at how to identify the key information intermediaries, another promising angle is the identification of individuals holding highly specialised roles. Criminal organisations are increasingly infiltrating highly specialised activities that require very specific knowledge, skills and competences. For example, pharmacology and chemistry expertise are required for synthetic drug manufacturing processes. The removal of these highly specialised nodes could decisively undermine the resilience of criminal organisations, as such individuals may be extremely difficult to replace.

This analysis paves the way to a vast range of further analyses. What emerges is that conventional analysis based on node centrality are insufficient on their own. New metrics have to be considered to gain better insights into Mafia clans interconnections, which communicate differently from other social networks. One possibility would be to combine popular centrality metrics with new ones that better capture these anomalous types of communications. The clan bosses are indeed the most powerful individuals. Yet they appear to generate the least frequent interactions. Instead, the soldiers (or "picciotti") emerge as the most important nodes. Thus, conventional network analysis will fail in identifying the bosses.

Under this prospective, it may also be possible to analyse the role that small traders (*e.g.* greengrocers or bakers) have in facilitating the interactions within and outside the Families mobsters. Observing how often a member of a clan meets people outside the organisation, could make it possible to discover communication

patterns and, in turn, differentiate between two types of interactions: (1) unrelated to the Mafia context (*e.g.* mobsters who occasionally buy something); (2) requests for protection (*i.e.* "pizzo" / racket) by the traders, which is typically periodical. This could be achieved using temporal networks analysis.

The next chapter will still relate on criminal networks, however the context is slightly different. We will address one of the biggest problems that LEAs face in their investigation, namely the missing data threshold for incomplete criminal graphs. Thus, the network features under scrutiny will be the network similarity metrics.

Chapter 4

Criminal Networks Part II: Missing Data Analysis

The work included in this chapter has been published in the following paper: [18].

4.1 Introduction

A significant problem in the analysis of real–world criminal networks is that important information is often mixed with vast amounts of irrelevant or unreliable information. In Social Network Analysis (SNA), the identification of relevant information from a dataset is usually referred to as the problem of signal and noise. Data collected in criminal investigations may suffer from issues like: (i) incompleteness, due to the covert nature of criminal organisations; (ii) incorrectness, caused by either unintentional data collection errors or intentional deception by criminals; (iii) inconsistency, when the same information is collected into law enforcement databases multiple times, or in different formats.

Thus, in this chapter nine real criminal networks of different nature (*i.e.*, Mafia networks, criminal street gangs and terrorist organisations) are analysed to quantify the impact of incomplete data, and to determine which network type is most affected by it. The networks were firstly pruned using two specific methods: (i) random edges removal, simulating the scenario in which the Law Enforcement Agencies (LEAs) fail to intercept some calls, or to spot sporadic meetings among suspects; (ii) nodes removal, modelling the situation in which some suspects cannot be intercepted or investigated. Finally, we computed spectral distances (*i.e.*,

Adjacency, Laplacian and normalised Laplacian Spectral Distances) and matrix distances (*i.e.*, Root Euclidean Distance) between the complete and pruned networks, which we compared using statistical analysis. Our investigation identifies two main features: first, the overall understanding of the criminal networks remains high even with incomplete data on criminal interactions (*i.e.*, when 10% of edges are removed); second, removing even a small fraction of suspects not investigated (*i.e.*, 2% of nodes are removed) may lead to significant misinterpretation of the overall network.

4.2 Related Works

Over the last decades, SNA has been employed greatly by LEAs. This increasing interest from law enforcement is due to the SNA ability to identify mechanisms that are not easily discovered at a first glance [108].

SNA relies on real datasets used as sources which allow to build networks that are then examined [86, 128, 129, 126, 16, 81, 17, 130]. However, the collection of complete network data describing the structure and activities of a criminal organisation is difficult to obtain.

In this work, a network science approach is adopted to assess how much of the available data of a criminal network may be missing, before it starts to be unreliable. In other words, our aim is to quantify how much the partial knowledge of a criminal network can affect investigations in a significant way.

An interesting application of SNA consists of comparing networks, by finding and quantifying similarities and differences between them [131, 132, 133]. Network comparison requires measures for the distance between graphs, a non-trivial task involving sets of features which are often sensitive to the specific application domain. Some reviews on the most common graph comparison metrics are [134, 135, 136, 137]. In an our earlier publication [138], such distance measures were exploited to quantify how well artificial (but realistic) models can simulate real criminal networks. The same measures are used herein for a different task. In this chapter, we analyse nine real criminal networks of different nature, which are the result of different investigative operations over Mafia networks, criminal street gangs and terrorist organisations. To quantify the impact of incomplete data and to determine what kind of network mostly suffers from it, we adopted the following strategy: (i) We pruned input networks by means of two specific methods, namely: *random edge removal* and *random node removal*, which reflect

the most common scenarios of missing data arising in investigation environments. (ii) We calculated the distance between the original (and complete) network and its pruned version.

4.3 Materials and Methods

This section presents the datasets used in our experimental analysis (Sect. 4.3.1) and provides a preliminary analysis of the degree distribution of the respective graphs obtained (Sect. 4.3.2). Finally, the protocol followed to run our analysis will be summarised in Sect. 4.3.3.

Useful details about Mafia, street gangs and terrorist networks are provided in Tables 4.1 and 4.2, including edges weight and directionality, connectedness, number of nodes including isolated ones, number of edges, number of connected components, maximum average path length for each connected component, maximum shortest path length, average degree, maximum degree and the average clustering coefficient. The CV network seems to be the only fully connected network (*i.e.*, $|cc| = 1$) and, for this reason, in all the considered networks we chose to compute the average path length for the single components and then to show the maximum value.

Table 4.1: **Mafia networks properties.**

Network weights	**MN** weighted	**PC** weighted	**SN** weighted	**WR** unweighted	**AW** unweighted	**JU** unweighted
directionality	undirected	undirected	undirected	undirected	undirected	undirected
connectedness	false	false	false	false	false	false
n. of nodes n	101	100	156	182	182	182
n. of isolated nodes n_i	0	0	5	0	36	93
n. of edges m	256	124	1619	247	189	113
n. of components $\|cc\|$	5	5	6	3	38	96
max avg. path length $\langle d \rangle$ for cc	3.309	3.378	2.361	3.999	4.426	3.722
max shortest path length d	7	7	5	8	9	7
density δ	0.051	0.025	0.134	0.015	0.011	0.007
avg. degree $\langle k \rangle$	5.07	2.48	20.76	2.71	2.08	1.24
max degree k	24	25	75	32	29	13
avg. clust. coeff. $\langle C \rangle$	0.656	0.105	0.795	0.149	0.122	0.059

4.3.1 Datasets Collection

As previously asserted, our analysis focuses on nine real criminal networks of different nature (see Table 4.3). In detail, the first six networks relate to three distinct Mafia operations, while the other three are linked to street gangs and terrorist organisations.

Table 4.2: **Street gangs and terrorist networks properties.**

Network weights	**SV** weighted	**CV** weighted	**PK** weighted
directionality	undirected	undirected	undirected
connectedness	false	true	false
nr. of nodes n	234	110	246
nr. of isolated nodes n_i	12	0	16
nr. of edges m	315	205	2571
nr. of components cc	13	1	26
max avg. path length $\langle d \rangle$ for cc	3.534	2.655	3.034
max shortest path length d	6	5	9
density δ	0.012	0.034	0.085
avg. degree $\langle k \rangle$	2.69	3.73	20.9
max degree k	34	60	78
avg. clustering coeff. $\langle C \rangle$	0.15	0.335	0.753

Table 4.3: **Criminal networks characterisation.**

Investigation	**Network**			**Source**
	Name	**Nodes**	**Edges**	
Montagna Operation (Sicilian Mafia) 2003-2007	MN PC	Suspects	Physical Surveillance Audio Surveillance	[16, 81, 17, 138, 125]
Infinito Operation (Lombardian 'Ndrangheta) 2007-2009	SN	Suspects	Physical and Audio Surveillance	[139, 140, 141, 142, 143]
Oversize Operation (Calabrian 'Ndrangheta) 2000-2009	WR AW JU	Suspects	Audio Surveillance Physical Surveillance Audio Surveillance	[89, 144]
Swedish Police Operation (Stockholm Street Gangs) 2000-2009	SV	Gang members	Physical Surveillance	[128, 145]
Caviar Project (Montreal Drug Traffickers) 1994-1996	CV	Criminals	Audio Surveillance	[1]
Abu Sayyaf Group (Philippines Kidnappers) 1991-2011	PK	Kidnappers	Attacks locations	[146]

Montagna Operation: The Montagna Operation was an investigation concluded in 2007 by the Public Prosecutor's Office of Messina (Sicily) focused on the Sicilian Mafia groups known as Mistretta and Batanesi clans from which we obtained the Meeting network (MN) and the Phone Calls network (PC) [16, 81, 17, 138, 130, 147, 125] (extensively discussed in Sect. 3.3.1).

Infinito Operation: The Infinito Operation was a large law enforcement operation against 'Ndrangheta groups (*i.e.*, groups of the Calabrian Mafia) and Milan cosche (*i.e.*, crime families or clans) concluded by the courts of Milan and Reggio Calabria, Italian cities situated in Northern and Southern Italy, respectively.

The investigation started 2003 is still in progress. On July 5, 2010, the Preliminary Investigations Judge of Milan issued a pre-trial detention order for 154 people, with charges ranging from mafia-style association to arms trafficking, extortion and intimidation for the awarding of contracts or electoral preferences. The dataset was extracted from this judicial act and is available as a 2-mode matrix on the UCINET [148] website[1]. The Infinito Operation dataset was investigated by Calderoni and his co-authors in several works [139, 140, 141, 142, 143]. From the original 2-mode matrix, we constructed the weighted and undirected graph Summits Network (SN) with 156 nodes and 1619 edges (Table 4.1). Nodes are suspected members of the 'Ndrangheta criminal organisation. Edges are summits (*i.e.*, meetings whose purpose is to make important decisions and/or affiliations, but also to solve internal problems and to establish roles and powers) taking place between 2007 and 2009. This network describes how many summits any two suspects may have in common. Attendance at summits was registered by police authorities through wiretapping and observations during this operation.

Oversize Operation: The Oversize Operation is an investigation lasting from 2000 to 2006, which targeted more than 50 suspects of the Calabrian 'Ndrangheta involved in international drug trafficking, homicides, and robberies. The trial led to the conviction of the main suspects from 5 to 22 years of imprisonment between 2007-2009. Berlusconi et al. [89] studied three unweighted and undirected networks extracted from three judicial documents corresponding to three different stages of the criminal proceedings (Table 4.1): wiretap records (WR), arrest warrant (AW), and judgement (JU). Each of these networks has 182 nodes corresponding to the individuals involved in illicit activities. The WR network has 247 edges that represent the wiretap conversations transcribed by the police and considered relevant at first glance. The AW network contains 189 edges which are meetings emerging from the physical surveillance. The JU network has 113 edges which are wiretap conversations emerging from the trial and several other sources of evidence, including wiretapping and audio surveillance. These datasets are available as three 1-mode matrices on Figshare [144].

Swedish Police Operation: The Stockholm street gangs dataset was extracted from the National Swedish Police Intelligence (NSPI), which collects and registers the information from different kinds of intelligence sources to identify gang membership in Sweden. The organisation investigated here is a Stockholm-based street

[1] https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/ndranghetamafia2

gang localised in southern parts of Stockholm County, consisting of marginalised suburbs of the capital. All gang members are male with high levels of violence, thefts, robbery and drug-related crimes. Rostami and Mondani [128] constructed the Surveillance (SV) network (Table 4.2). It contains data from the General Surveillance Register (GSR) that covers the period 1995–2010 and aims to facilitate access to the personal information revealed in law enforcement activities needed in police operations. SV is a weighted network with 234 nodes that are gang members. Some of them were no longer part of the gang in the period covered by the data and have been included as isolated nodes. The link weight counts the number of occurrence of a given edge. This dataset is available on Figshare [145].

Project Caviar: Project Caviar [1] was a unique investigation against hashish and cocaine importers operating out of Montreal, Canada. The network was targeted between 1994 and 1996 by a tandem investigation uniting the Montreal Police, the Royal Canadian Mounted Police, and other national and regional law-enforcement agencies from England, Spain, Italy, Brazil, Paraguay, and Colombia. In a 2-year period, 11 imported drug consignments were seized at different moments and arrests only took place at the end of the investigation. The principal data sources are the transcripts of electronically intercepted telephone conversations between suspects submitted as evidence during the trials of 22 individuals. Initially, 318 individuals were extracted because of their appearance in the surveillance data. From this pool, 208 individuals were not implicated in the trafficking operations. Most were simply named during the many transcripts of conversations, but never detected. Others who were detected had no clear participatory role within the network (*e.g.*, family members or legitimate entrepreneurs). The final Caviar (CV) network was composed of 110 nodes. The 1-mode matrix with weighted and directed edges is available on the UCINET [148] website[2]. From this matrix, we extracted an undirected and weighted network with 110 nodes that are criminals and 205 edges representing the communications exchanges between them (see Table 4.2). Weights are level of communication activity.

Abu Sayyaf Group: Philippines Kidnappers data refer to the Abu Sayyaf Group (ASG) [146], a violent non-state actor operating in the Southern Philippines. In particular, this dataset is related to the Salast movement that has been founded by Aburajak Janjalani, a native terrorist of the Southern Philippines in 1991. ASG is active in kidnapping and other kinds of terrorist attacks. The recon-

[2] https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/caviar

structed 2-mode matrix is available on UCINET [148][3]. From the 2-mode matrix, we constructed a weighted and undirected graph called Philippines Kidnappers (PK) (see Table 4.2). The PK network has 246 nodes and 2571 edges. Nodes are terrorist kidnappers of the ASG. Edges are the terrorist events they have attended. This network describes how many events any two kidnappers have in common.

4.3.2 Dataset Statistics

Herein, the degree distributions for each criminal network as a normalised histogram (see Fig. 4.1) is showed. MN, PC, WR, AW, JU, SV and CV have similar degree distributions in which most nodes have a relatively small degree k with values around 0, 1 or 2, while a few nodes have very large degree k and are connected to many other nodes. SN and PK are the only networks having different degree distributions compared to other criminal networks, as most of their nodes have large degree k. In particular, we note that most nodes in PK are strongly connected and have a degree $k = 57$.

SN, which derives from the Infinito operation, is a one-mode projection of the original two-mode network in which are represented the meetings and the suspects attending them. This implies that all suspects taking part in a meeting are assumed to be interacting with each other, which could be somewhat artificial. In fact, in crowded meetings some participants may have had a very limited (if any) interaction with other participants. In such case, assuming that all participants interacted with each other may considerably overestimate the real number of connections. However, it must be added that LEAs were only able to identify the participants to meetings and not the full extent of their interactions. Similar consideration applies to PK, which was built based on the presence of the kidnappers in the same place of a terrorist event. Here as well, the existence of an edge linking two terrorists does not necessarily imply that they have interacted or worked together, despite being in the same place.

4.3.3 Design of Experiments

In this section we give technical details on the design of the experiments conducted.

In the attempt of gaining a deeper understanding of criminal networks, in our previous work [138] we used graph distances to compare randomly generated graphs

[3] https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/philippinekidnappings

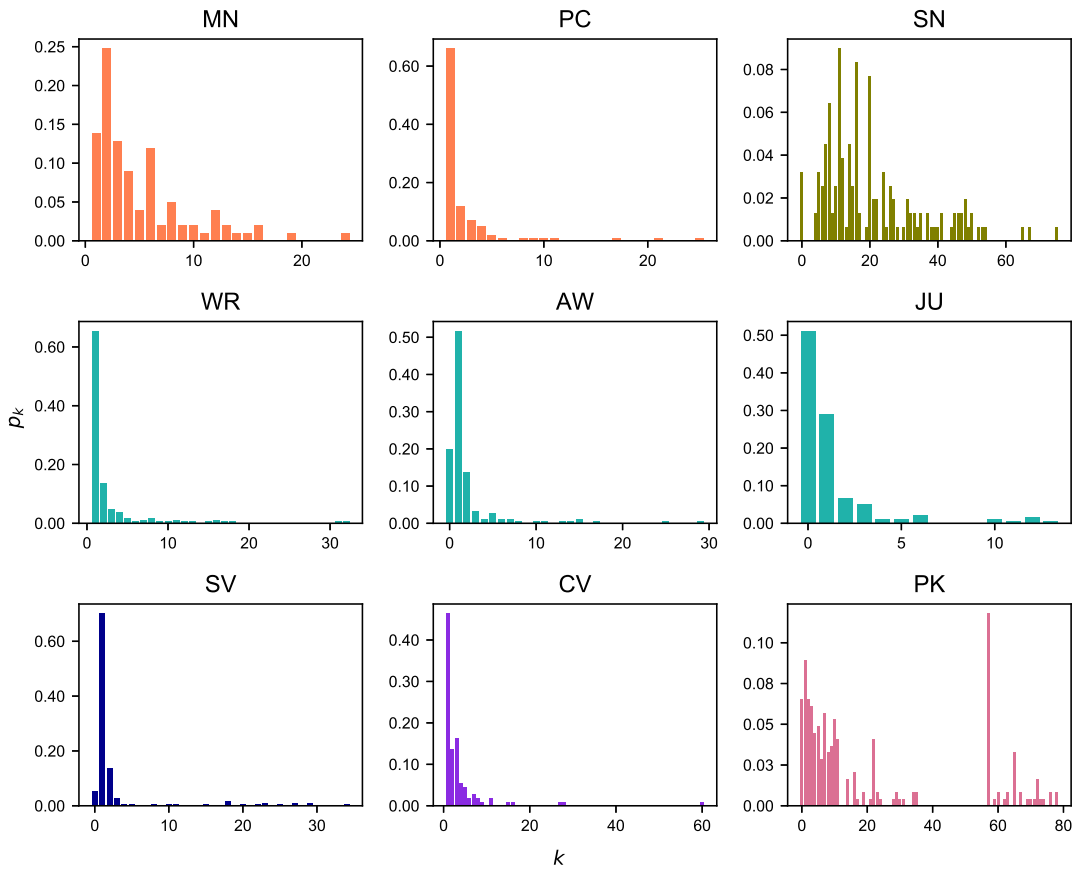


Figure 4.1: **Degree Distributions.** The degree distribution p_k provides the probability that a randomly selected node in each criminal network has degree k. Same colors imply the networks belong to the same police investigation.

and a real criminal network. In the present work, we have implemented distances to understand to which extent a partial knowledge of a criminal network may negatively affect the investigations. Since we are trying to estimate differences based on the types/amount of missing data, we set up the experiments based on two main strategies: random edges removal and nodes removal. The first case simulates the scenario in which LEAs miss to intercept some calls or to spot sporadic meetings among suspects (*i.e.*, due to the delays in obtaining a warrant). In nodes removal, the selected nodes are removed along with their incident edges, and afterwards they are reinserted within the networks as isolated nodes. Indeed, the second case models the scenario in which some suspects cannot be intercepted. For instance, if a criminal is known to be a boss but there are not enough proofs to be investigated, then that criminal can be identified as an isolated node with no incident edges. However, node removal is expected to have a greater impact than simple edge removal, since removing a node implies the systematic deletion of all its edges as well.

Note that for a better comparison among the networks, all the graphs have been considered as unweighted (as AW and JU are). Also, all the suspects showed as isolated nodes of the original network have been excluded. In fact, our input parameter was the edge list of the graph, which does not take into account nodes with no incident edges.

Algorithm 1 shows the pseudocode of our approach. The full code is available on GitHub[4]. In order to obtain the subgraphs, we started from the previously described datasets; then, we converted them into graphs (*i.e.*, G) and, lastly, we pruned them (*i.e.*, $G^{'}$) according to a prefixed range of fractions with $0 < torem \leq 10\%$. We opted for the 10% because the criminal networks considered are small, as they have less than 250 nodes. Afterwards, we have computed the spectral and matrix distances between the original and the pruned graphs. Each edges removal process has been repeated a fixed number of times ($nrep = 100$) and the results obtained have been averaged. Thus, the averaged distances values $\langle X \rangle$ and their standard deviations σ have been computed.

Algorithm 1 Pseudocode for computing the distances.

```
Parameter configuration: nrep, torem, and check
Read the dataset and convert it as graph G
if check = True then
    Isolate torem of nodes
else
    Remove torem of random edges
Compute S(G)
Compute the matrices A(G), L(G), Ł(G)
for torem do
    for nrep do
        Create a pruned graph G' and compute S'(G')
        Compute d_rootED(G, G'), d_A(G, G'), d_L(G, G'), and d_Ł(G, G')
    Compute ⟨X⟩, σ ∀ d(G, G') ∈ nrep
```

4.4 Missing Data Analysis Results

Herein, the results obtained from the network pruning experiments are presented. The distance analysis between the real and the pruned networks is reported starting from the random edges removal approach (Fig. 4.2), moving to the analysis on the networks after node pruning (Fig. 4.3). The plots show the distances between the original graphs and their pruned versions up to 10% of edges (F_e) and nodes

[4] https://github.com/lcucav/criminal-nets/tree/master/missing_data

(F_n), respectively.

In both removal processes, d_A displays a saturation effect that makes the results difficult to be interpreted. Hence, this distance is not effective for highlighting the effects of missing data on criminal networks. Furthermore, from this metric it might seem that the two pruned networks of PK and SN show a greater deviation from their original counterparts, but this is due to the inner structure of this metric, which is highly influenced by the nodes' degree. In fact, the average degree of PK an SN (see Tables 4.1 and 4.2) is significantly higher (*i.e.*, $\langle k \rangle \simeq 21$) than the other networks herein studied (*i.e.*, $1 < \langle k \rangle < 4$). Moreover, their different topology is also evident from their degree distribution (see Fig. 4.1). This is the reason why these networks seem to have a more significant detachment effect than others; however, they too suffer the saturation effect mentioned above as they grow. A similar behaviour has also been encountered in d_L and its explanation is the same.

On the other hand, the distance metric which more effectively catches the damage caused by a significant amount of missing data is $d_{\mathcal{L}}$, where distance growth is linear. Indeed, the effects of $\langle k \rangle$ are smaller as this aspect is compressed by the structure of this distance metric. It would seem that this metric is the most effective measure compared to other spectral distances, in understanding how much lacking data affects the total knowledge of the network. A similar trend was also found in $\mathrm{d_{rootED}}$; however, for a better comparison between nodes and edges removal processes, we analysed in more detail this last metric by considering the DeltaCon similarity sim_{DC} (Fig. 4.4).

The figure shows the difference between the original and pruned networks as the fraction of elements removed increases (*i.e.*, F_e for edges and F_n for nodes).

Before pruning the networks we have $sim_{DC} = 1$. Afterwards, the drop begins to became more evident as the fraction F increases. In addition, as expected, the nodes removal process affects more significantly the networks. This means that if the lack of data relates to sporadically missed wiretaps, or to just a few random connections between suspects, then the network structure is not as much misinterpreted as if the case when one suspect has not been tracked at all. Indeed, pruning the network by 2%, causes a $sim_{DC} \geq 0.8$ for edges pruning, compared to a $sim_{DC} \simeq 0.2$ for the nodes ones. Therefore, even when a small amount of suspects are not included in the investigations, this can lead to a very different network. The exclusion of the suspects could be voluntary or not. It highly depends on the overall investigation process, starting from the very preliminary

analysis, and up to the judges' decision to allow warrants, or to exclude data considered irrelevant for the current investigation.

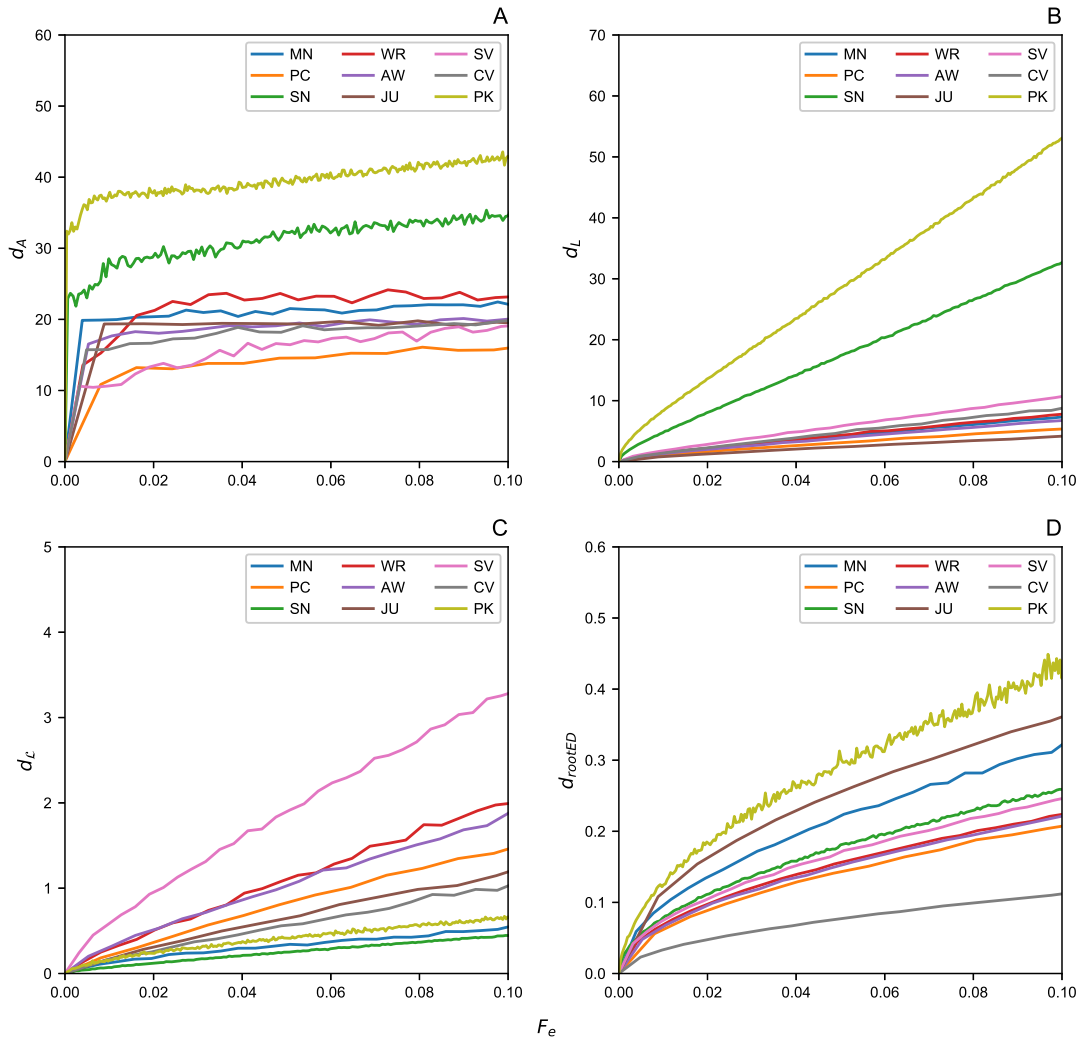


Figure 4.2: **Edges removal effect.** The removal effects of a fraction F_e of edges by showing the graph distances between the original graphs with their pruned versions. (A) Adjacency Spectral Distance d_A. (B) Laplacian Spectral Distance d_L. (C) Normalised Laplacian Spectral Distance $d_{\mathcal{L}}$. (D) Root Euclidean Distance $\mathrm{d}_{\mathrm{rootED}}$.

4.5 Conclusions

In this chapter, nine datasets of real criminal networks extracted from six police operations have been analysed with the aim of investigates the effects of missing data. More specifically, six datasets (*i.e.*, MN, PC, SN, WR, AW, Ju) regard Mafia operations (*i.e.*, Montagna, Infinito, and Oversize), and the remaining ones (SV, CV, PK) refer to other criminal networks, including street gangs, drug traffics,

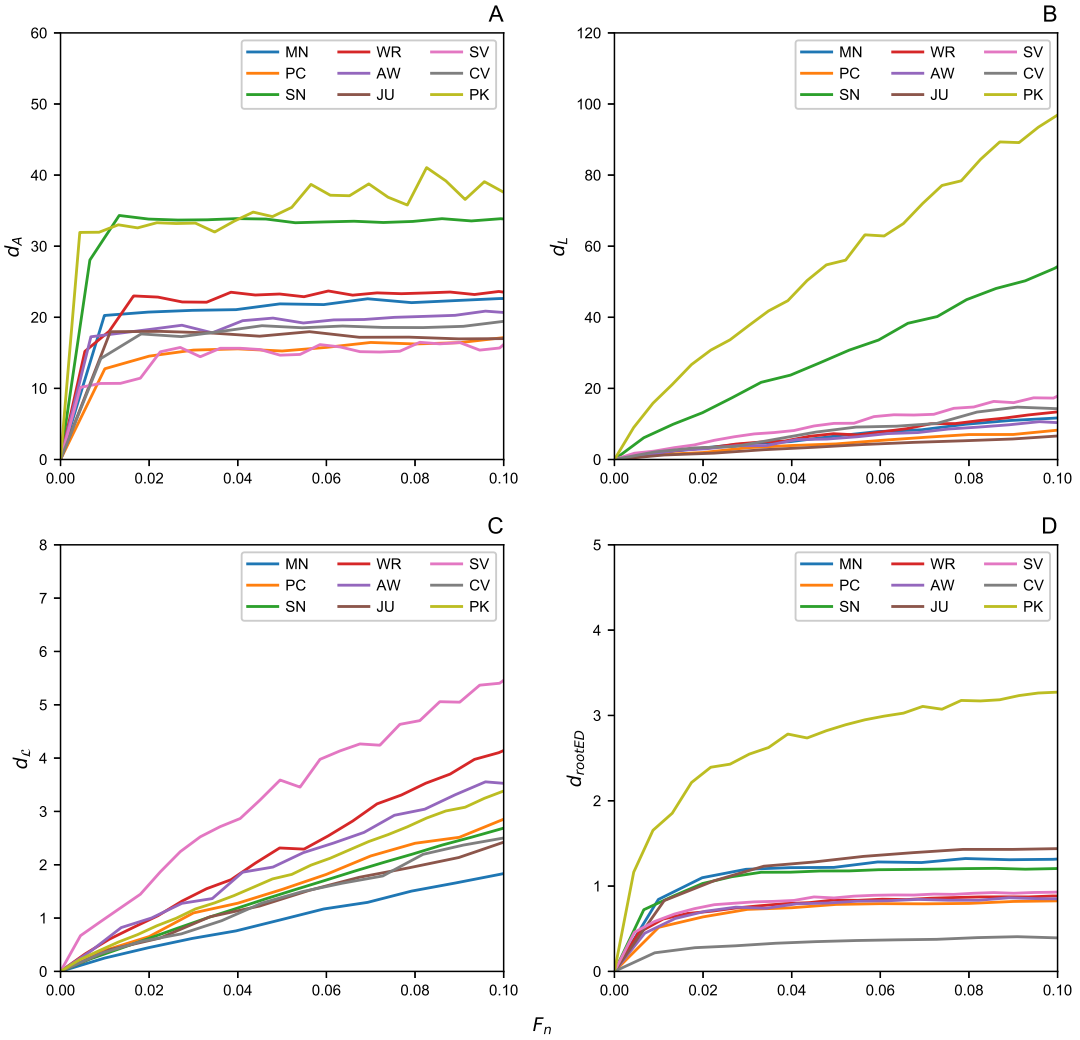


Figure 4.3: **Nodes removal effect.** The removal effects of a fraction F_n of nodes by showing the graph distances between the original graphs with their pruned versions. (A) Adjacency Spectral Distance d_A. (B) Laplacian Spectral Distance d_L. (C) Normalised Laplacian Spectral Distance $d_{\mathcal{L}}$. (D) Root Euclidean Distance d_{rootED}.

or terrorist networks (*i.e.*, Stockholm street gangs, Caviar Project, Philippines Kidnappers).

Our work focused on a careful analysis of the datasets, in order to simulate the events where some data are missing. In particular, two different scenarios have been considered: (i) random edges removal, simulating the case in which LEAs miss to intercept some calls or to spot sporadic meetings among suspects; and (ii) nodes removal, for the scenario where certain suspects cannot be intercepted for some reason. For instance, if a criminal is known to be a boss, but there are not enough proofs for him or her to be investigated, then this can be identified by an isolated node with no incident edges.

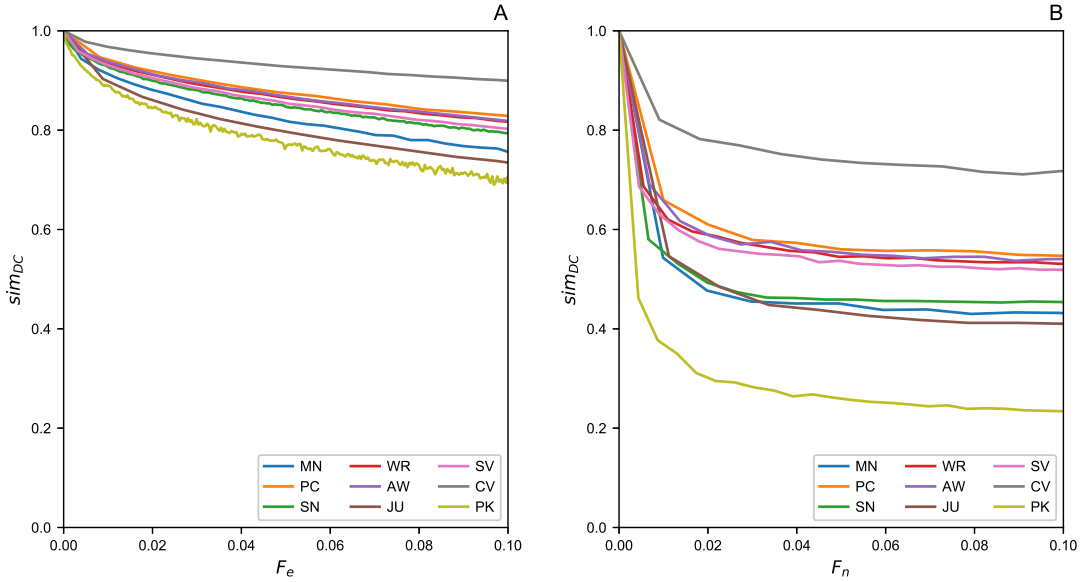


Figure 4.4: **DeltaCon similarity sim_{DC} computation**. (A) Edges removal process by the fraction F_e. (B) Nodes removal process by the fraction F_n.

To quantify the difference between the original criminal networks and their pruned counterparts, several distance metrics have been considered. We computed the Adjacency, Laplacian, and normalised Laplacian Spectral distances (*i.e.*, d_A, d_L, and $d_{\mathcal{L}}$, respectively) plus the Root Euclidean Distance (*i.e.*, $\mathrm{d_{rootED}}$), as this metric allows to compute the DeltaCon similarity (i,e., sim_{DC}), which can quantify even small differences between two graphs in the interval $[0, 1]$. The pruning process involved removing a fraction of up to 10% of edges and nodes. This percentage has been chosen as the networks size was quite small (less than 250 nodes per each dataset).

Our analysis suggests that (i) the spectral metric $d_{\mathcal{L}}$ is best at catching the expected linear growth of differences with the incomplete graph against its complete counterpart; (ii) the nodes removal process is significantly more damaging than random edges removal; thus, it translates to a negligible error in terms of graph analysis when, for example, some wiretaps are missing. Indeed, in terms of sim_{DC} drop, there is a 30% difference from the real network, for a pruned version at 10%. On the other hand, it is crucial to be able to investigate the suspects in a timely fashion, since any exclusion of suspects from an investigation may lead to significant errors (due to substantial differences from the actual network) - we observed drops of up to 80% of sim_{DC} on some networks.

A final consideration concerns the impossibility of conducting this type of analysis through the use of Machine Learning, as it is currently practically impossible

to obtain a sufficient number of reliable and complete datasets of real criminal networks as to be able to conduct an appropriate training of a Neural Network.

For the future, it could be interesting to conduct a similar analysis by considering weights as well. This will allow to conduct a comparative analysis of the missing data effects when not only the connections between nodes, but also their frequency is known. Another aspect to be considered is the network behaviour after their pruning in both criminal and general social networks. Lastly, using the future knowledge gained from the network analysis herein presented, one could try to define an artificial network able to accurately simulate the behaviour of real criminal networks.

The next chapter will cover another promising aspect of the use of Network Science, that is network robustness. Indeed, the network feature under scrutiny is the probability of a node to reject the attempt of being removed from a network. This may happen because of (i) an external attack or (ii) a node failure. The novelty introduced through this work is that in the state-of-the-art the nodes removal process is assumed to be always successful. By contrast, we introduce the probabilistic failure model that covers this gap.

Chapter 5

Probabilistic Failure Model

The work included in this chapter has not been published yet. While it is undergoing peer-review in IEEE Transactions on Network Science and Engineering, it is accessible in ArXiv [19].

5.1 Introduction

When dealing with complex networks, node-removal processes are crucial Networks Science tools, which may be used to test the network strength and verify its resilience. Node-removal processes are generally used to test network robustness against failures, verify the strength of a power grid, or contain fake news. The most significant gap in the state-of-art in this application domain is the assumption that the node-removal processes are deterministic (see Sect. 5.2). That is to say, that any attempt at removing a node is assumed to be *always successful* [30, 149, 5, 4, 150]. By contrast, we argue that this is unrealistic, and that the node strengths should also be considered to better accommodate network failure scenarios.

Indeed, if the goal were to analyse a well-engineered network, such as a power grid, we should expect nodes to exhibit some level of resilience, and to have been designed with a certain ability to resist to external attacks and failures. For instance, power grid nodes are equipped with means to resist to hacking and to hardware or software failures [4]. By contrast, there are networks, such as the social networks, that respond more deterministically during to node removal (*i.e.*, when a node is picked for removal, it will always be pruned). For instance, as it

has been shown in the previous experimental chapters so far, an example of social network is a criminal organisation network, like the Sicilian Mafia. In that context, criminals (*i.e.,* nodes) can certainly resist to an external attack (*i.e.*, probabilistic failure) by Law Enforcement Agencies (LEAs, *i.e.*, node removal). The resilience can be seen with the view to hidden their trails and making difficult from LEAs to collect enough proves to arrest them. However, as it has been extensively discussed previously in this thesis, it is difficult to define a criterion that takes into account both social and human capital to unequivocally detect the suspects. Indeed, as will be later shown in this chapter, in human engineering networks (*e.g.*, power grid networks) that have been built to be resilient, it is easier to compare our approach with the classic one (*i.e.*, which can be used as benchmark).

Thus, moving from a graph-theoretical approach to real–world network analysis requires one to question how the nodes react to node-removal, whether or not they comply, and the extent by which they are engineered to resist attacks and failure.

This challenge has lead us to conceive a new model to take into account the nodes failure profile. In this way, we can more realistically estimate how network connectivity (in a graph G) is actually affected by node removal, considering different types of networks (from low to high resilience), nodes' role, and centrality metrics.

To this end, this chapter introduces a *probabilistic node failure model* $\mathcal{M}$, which associates each node with its probability to survive a failure. When the node survival probability is zero (herein addressed as "Benchmark" analysis), the model coincides with other models already introduced in the literature [4, 5].

We considered two variants of our model: (i) *Uniform* nodes survival-to-failure probability; and (ii) *Best Connected* (survival probability proportional to node degree). Our evaluation considers five popular centrality metrics (degree, h-index, coreness, Eigenvector, Katz centrality), performing an experimental analysis on *effectiveness* and *coverage*, on eight real–world graphs. Specifically, the effectiveness is defined as the drop in the spectral radius λ_1 after node removal, while the coverage is understood as the reduction of the size of the largest connected component c of a graph. Note that this last evaluation metric has been already used in the context of criminal networks in Chapt. 3. As presented next, we found that the node degree can generally be used to cause the biggest drop in both λ_1 and c, especially in graphs deriving from human interactions/collaborations. Comparing with conventional methods, our probabilistic model exhibits significant differences (ranging from 0% to 83%), highlighting the benefits of our method.

5.2 Related Works

This section reviews past research works related to this thesis. Firstly, the ability of centrality metrics to identify nodes in a graph that give rise and favour diffusion processes is addressed (Section 5.2.1). Then, in Section 5.2.2, approaches that manipulate graph topology are reviewed, exploring how these modifications alter centrality metrics, as well as other graph parameters.

5.2.1 Identifying nodes capable of activating diffusion processes

The problem of detecting nodes that originate diffusion processes in networks has been extensively studied in the past, and is well aligned with the problem of calculating centrality scores in graphs [149, 151]. A relevant application is the study of the misinformation spreading in OSNs [152, 12].

For instance, Comin and da Fontoura Costa [152] applied standard centrality metrics like degree centrality to identify the sources of misinformation. Shah and Zaman [153] introduced an ad-hoc centrality parameter, called *rumour centrality*, to rank nodes on the basis of their spreading ability. They focused on tree-like networks and hypothesised that a node can receive information from only one of its neighbours. Dong *et al.* [154] extended the approach of Shah and Zaman [153] by detecting nodes with the largest rumour centrality within a set of suspected nodes.

Contrary to the aforementioned approaches, we consider graphs of arbitrary topology and we are on a quest to assess how the spectral radius λ_1 and the LCC size c of a graph vary upon the random failure of some nodes.

Prakash *et al.* [155] applied spectral methods to evaluate the spreading power of nodes. However, due to its high computational costs, their method is applicable only to small-size graphs. Nguyen *et al.* [156] employed Monte Carlo techniques to discover the set of nodes that are the best candidates for spreading misinformation. Budak *et al.* [157] considered competing campaigns over a social network and aimed at identifying a subset of individuals that need to be convinced to promote a "good" campaign to minimise the number of people who adhere to a "bad" one. They proved that degree centrality is a good heuristic to find out nodes involved in good campaigns, provided that the delay elapsing between the start of misinformation spread and its first detection is fairly small.

These methods assume that some nodes in networks should be better protected to neutralise misinformation spread. Such a belief agrees with a core assumption of our approach: in fact, in the Best Connected (BC) model, we suppose that large-degree nodes occupy a crucial role in the system functioning. Thus, they must display a larger survival probability to failures.

Coming shortly back to the criminal networks context, even though the bosses tend to minimise their interaction, the assumption that nodes with large degree have a crucial role in spreading information remains still valid. As previously asserted, more specifically, the betweenness centrality metric has been revealed to be the most effective one when ranking the suspect importance (within the criminal organisation) in connecting the otherwise isolated parts of the network. For the sake of computation simplicity, herein we have considered the BC model taking into account node degree.

A relevant difference between approaches from the literature and ours is that their neutralisation strategy requires to solve an optimisation problem; whereas we are in charge of evaluating the deformation of λ_1 and c. In addition, as a guiding criterion for the selection of nodes to be removed, we adopt centrality metrics, which are easy to calculate and have a clear interpretation.

5.2.2 Variation in graph connectivity after node removal

Node removal procedures have been applied to investigate the resilience of large systems [4, 158, 57]. Albert *et al.* [4] studied how the diameter and the size of the giant component of Erdős-Rényi and scale-free graphs varied when nodes were removed at random, or if large-degree nodes were deleted. Borgatti *et al.* [158] examined the accuracy in estimating centrality scores when graph data are incomplete. In the Web search domain, Ng *et al.* [159] examined small changes of the Web graph and their impact on the PageRank and HITS scores.

Restrepo *et al.* [57] defined the *dynamical importance* of a node i as the amount $-\Delta\lambda_i$ by which λ_1 decreases upon the removal of the all edges incident onto i, normalised by λ_1. Al-Dabbagh [160] studied the topology design of a wireless control system in which nodes and wireless links are unreliable (for instance, due to battery drainage). The approach of [160] (shared by us) assumes that network elements may fail in an unpredictable way; yet the focus is to determine whether it is possible to design a controller for a wireless network, given that the largest number of unreliable nodes in the network, and the probability that a link fails are specified.

Unlike approaches described in literature, we manage a scenario in which nodes may survive a failure. Thus, we considered a probabilistic framework to describe node failure. In addition, many of the approaches discussed in this section fix the number k of nodes/edges to be removed, and attempt to find the best strategy to delete at most k nodes. In contrast, we aim at experimentally studying the variation in λ_1 and c when the fraction of nodes subject to failure varies.

5.3 Materials and Methods

This section firstly shows the datasets we used for our analyses (Sect. 5.3.1) followed by a description of the two variants of the probabilistic failure model we developed, namely: *Uniform*, in which the probability that any node survives a failure is equal to a fixed value p; and *Best Connected* (BC), in which the probability that a node survives a failure is proportional to its degree.

Next, to quantify the loss of connectivity (and fragmentation) in G, for both the Uniform and BC models, in Sect. 5.3.2, we assess both the *spectral radius* λ_1 and the *largest connected component* (LCC) size c. As already discussed in Chapt. 2, the spectral radius is defined as the adjacency matrix[1] largest eigenvalue of G, and governs a broad range of spreading processes in G, such as the diffusion of an infection [6, 7, 8, 9], malware propagation [10, 11], or the dissemination of fake news in Online Social Networks (OSNs) [12, 13, 14]. On the other hand, the LCC size is defined as the number of nodes in the largest connected subgraph of G, and is widely used to quantify the resilience of a natural or artificial system described by the graph G [4] (as we did in Chapt. 3).

With those assumptions in mind, we finally simulate a *node removal process* in which nodes may fail according to either the Uniform or the BC model. Failed nodes are sorted, in turn, by the degree, h-index, coreness, Eigenvector, and Katz centrality scores. For brevity, ϕ herein denotes any of the centrality metrics above. Next, the top $\lceil \tau \times |N| \rceil$ nodes from the node ranking generated by ϕ are deleted, being N the set of nodes of G and τ a fixed threshold in $[0, 1]$. This process generates a graph $\tilde{G}$ with spectral radius $\tilde{\lambda}_1$ and the LCC size $\tilde{c}$.

Lastly, Sect. 5.3.2 will describe in detail the framework of our approach.

[1] The adjacency matrix $\mathbf{A}$ of a graph G is defined as $\mathbf{A}_{ij} = 1$ if nodes i and j are connected, 0 otherwise.

5.3.1 Datasets

Eight real-life graphs have been employed to perform the experimental analysis, each one taken from the SNAP repository[2] namely: FACEBOOK (composed by the circles from Facebook), US_POWER_GRID (describing the power grid in US Western states), LASTFM (a social network of LastFM users from Asian countries), CA-HEPPH (that is a collaboration network between papers' authors from High Energy Physics - Phenomenology category of ArXiv), ASTROPH (describing co-authorship between scientists in the astrophysics domain), ENRON (a communication networks where the nodes represent the email addresses), BRIGHTKITE (a location-based social networking Web site), and FLICKR (a graph whose nodes represent Flickr photos and an edge indicates that two photos share some tags). The datasets features are summarised in Table 5.1.

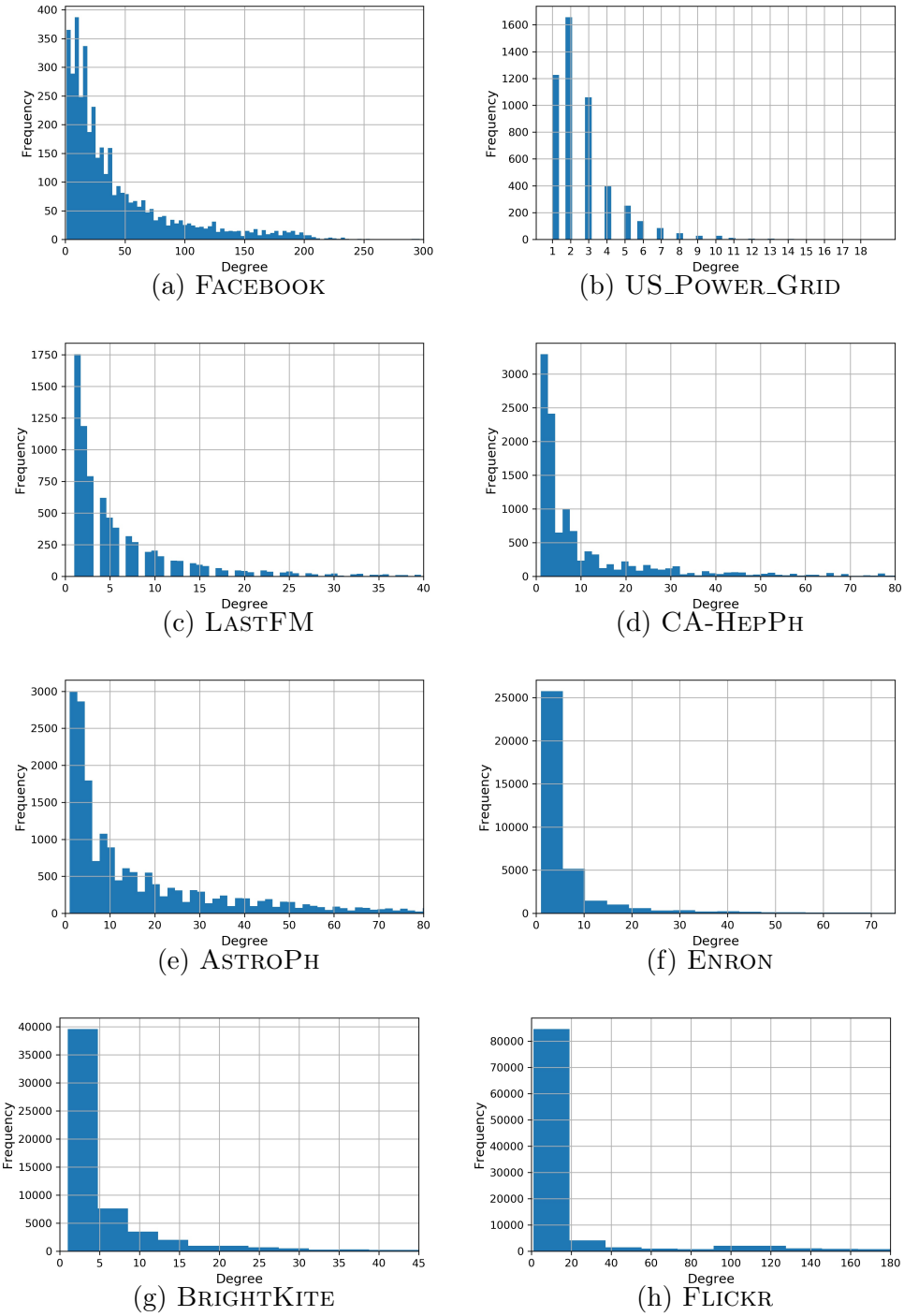
Table 5.1: Datasets adopted in the experimental trials. For each dataset the number of nodes, and the number of edges are reported.

	Dataset	# Nodes	# Edges	Ref
1	FACEBOOK	4 039	88 234	[161]
2	US_POWER_GRID	4 941	6 594	[39]
3	LASTFM	7 624	27 806	[162]
4	CA-HEPPH	12 008	118 521	[163]
5	ASTROPH	18 771	198 050	[163]
6	ENRON	36 692	183 831	[164]
7	BRIGHTKITE	58 228	214 078	[165]
8	FLICKR	105 938	2 316 948	[166]

In Figure 5.1 the degree distribution of all graphs involved in this study is reported.

The degree distribution in FACEBOOK dataset reports more than 4 000 users having around 10 contacts; similarly, most of the nodes in LASTFM display a degree less than three. In CA-HEP and in ENRON the degree distribution quickly vanishes if the degree is greater than 10 or 35, respectively. Nonetheless, the degree distribution in US_POWER_GRID significantly deviates from that observed in other graphs: herein, most of the nodes have a degree comprised from one to three, but a non-negligible fraction of nodes has a degree comprised from four to seven. Instead, for bigger datasets, such as in BRIGHTKITE and FLICKR, node degree distribution is highly skewed. In the first scenario, (Fig. 5.1(g)) roughly 70% of nodes have degree less than five, while in the second one (Fig. 5.1(h)) around one thousand nodes in have a degree from 90 to 130.

[2] http://snap.stanford.edu/data/index.html



(a) FACEBOOK (b) US_POWER_GRID

(c) LASTFM (d) CA-HEPPH

(e) ASTROPH (f) ENRON

(g) BRIGHTKITE (h) FLICKR

Figure 5.1: Degree Distribution in the input datasets.

5.3.2 A probabilistic node failure model

This section presents our probabilistic node failure model, and the proposed protocol aimed to analyse both node failure impact on the spectral radius λ_1 and the largest connected component (LCC) size c of a graph G. Let λ_1 (resp., c) be the spectral radius (resp., the LCC size of G).

Our methodology consists of the following components: (i) an undirected graph $G = \langle N, E \rangle$. (ii) A *node-scoring* function $\phi : N \to \mathbb{R}^+$, which takes a node i as input and returns its centrality $\phi(i)$ as output. Herein, the degree, h-index, coreness, Eigenvector, and Katz centrality are evaluated. (iii) A *survival probability* function $\psi : N \to [0, 1]$, which takes a node i as input and returns the probability $\psi(i)$ that i will survive a failure. (iv) A *target threshold* $\tau \in [0, 1]$, which specifies the fraction of nodes subject to failure.

Our methodology comprises the following steps:

1. Each node $i \in N$ is associated with a score $\sigma_i = (1 - \psi(i))\,\phi(i)$. Herein, high-score nodes are those with a high centrality (encoded in the factor $\phi(\cdot)$) and with a large probability of failing (expressed as $1 - \psi(\cdot)$).

2. The top $\lceil \tau |N_0| \rceil$ nodes with largest scores are picked and deleted from G along with their edges.

Two variants of *probabilistic node failure model*, namely *Uniform* and *Best Connected*(BC) are considered. In the *Uniform* model, it supposes that $\psi(i) = p$ for every node i, being p a fixed value in $[0, 1]$. The BC model is grounded on the principle that nodes display an unequal level of tolerance to failures: intuitively, large-degree nodes have to occupy a prominent position in G because their removal may quickly lead to network fragmentation [4]; thus, they should display a better resistance to failures. A possible model of $\psi(\cdot)$ – which incorporates the observations above – is $\psi(i) = \frac{d_i}{2m}$, where $d_i(\frac{2m-d_i}{2m}) \simeq d_i$ for large n. Other models to describe node resistance to failures are also allowed, but we leave their discussion as future work.

The procedure above yields a new graph $\tilde{G}(\tau, \phi)$ with spectral radius $\tilde{\lambda}_1(\tau, \phi)$ and the LCC size $\tilde{c}(\tau, \phi)$ and it is interesting to compare $\tilde{\lambda}_1(\tau, \phi)$ (resp., $\tilde{c}(\tau, \phi)$) with λ_1 (resp., c).
To this end, observe that $\tilde{\lambda}_1(\tau, \phi)$ is not greater than λ_1 (see [167] for a proof). Thus, we can compute the ratio – called *effectiveness* – between $\tilde{\lambda}_1(\tau, \phi)$ and λ_1 to quantify the drop in the spectral radius: the closer to zero this ratio, the more significant the reduction in the spectral radius.
In an analogous fashion, let $\tilde{c}(\tau, \phi)$ be the size of the LCC of $G(\tau, \phi)$. Nodes removal from G leads to a shrinkage in the LCC size in $\tilde{G}(\tau, \phi)$, which implies that $\tilde{c}(\tau, \phi)$ is no greater than c. Therefore, we can calculate the ratio – called *coverage* – of $\tilde{c}(\tau, \phi)$ to c to estimate the amount of reduction in the size of LCC and the closer such a ratio to zero, the higher the shrinkage of the largest connected

component.
A more formal definition of effectiveness and coverage is reported below:

Definition 1. *Let G be an undirected and connected graph and let $\tau \in [0,1]$. Let $\tilde{G}(\tau,\phi)$ be the graph obtained from G by applying the probabilistic node failure model above with ϕ as centrality metric and τ as target threshold.*

The effectiveness $\rho(\tau,\phi)$ of ϕ in the Uniform (resp., BC) model is defined as:

$$\rho(\tau,\phi) = \frac{\tilde{\lambda}_1(\tau,\phi)}{\lambda_1} \tag{5.1}$$

where $\tilde{\lambda}_1(\tau,\phi)$ (resp., λ_1) is the spectral radius of $\tilde{G}(\tau,\phi)$ (resp., G).

The coverage $\gamma(\tau,\phi)$ of ϕ in the Uniform (resp., BC) model is defined as:

$$\gamma(\tau,\phi) = \frac{\tilde{c}(\tau,\phi)}{c} \tag{5.2}$$

where $\tilde{c}(\tau,\phi)$ (resp., c) is the LCC size of $\tilde{G}(\tau,\phi)$ (resp., G).

In Section 5.4 the variation of effectiveness and coverage is analysed, through a set of experiments on real-life graphs.

5.3.3 Design of Experiments

In the experiments, the fraction τ of nodes to be removed has been varied from 0 to 0.18. As for the Uniform model, three different values of the survival probability p, namely, 0.1, 0.3, and 0.5 have been considered.

Note that in all the effectiveness and coverage plots (Figures 5.2-5.17), the Benchmark analysis has been reported too. This curve is the same in all the plots as it represents the normal behaviour of the node-removal process, when the probabilistic failure is not considered (*i.e.*, as if $p = 0$ in the Uniform model), by using the degree centrality metric, as it will be later confirmed to be the most effective one for the analysis of both λ_1 and c drops. Thus, the Benchmark allows to make a comparison between the state-of-art approach and our more realistic proposal.

An additional layer of complexity in running our experimental tests depends on the tuning of the α parameter in the computation of the Katz centrality.We recall that, if α approaches zero, then the node ranking generated by the Katz centrality

converges to that produced by the degree centrality; analogously, if α tends to its upper bound (namely $\frac{1}{\lambda_1}$), then the node ranking due to the application of the Katz centrality converge to that obtained by the Eigenvector centrality [50]. Therefore, we concentrated on values of α inside the interval $(0, \frac{1}{\lambda_1})$ that were reasonably far from the lower and the upper bounds. Previous studies ([50, 168]) pinpoint that the rankings produced by the Katz coefficient for different values of α are quite *stable*: that is, (even small) variations in α may imply that actual Katz coefficient *scores* vary of some orders of magnitude, but there are not significant changes in the node rankings deriving from different choices of α. Our approach consists in targeting the set composed by the top $\lceil \tau |N_0| \rceil$ nodes with the largest Katz centrality scores; thus, a preliminary experiment has been conducted to test how the set of nodes to target varies for different choices of α. Our tests confirmed the results presented in [50, 168]; *i.e.*, we did not report significant changes in the set of nodes to target if we choose different values of α. In the light of the considerations above, we fixed $\alpha = \frac{0.1}{\lambda_1}$.

The experiments' results have been averaged 10 times to avoid statistical fluctuations.

5.4 Results

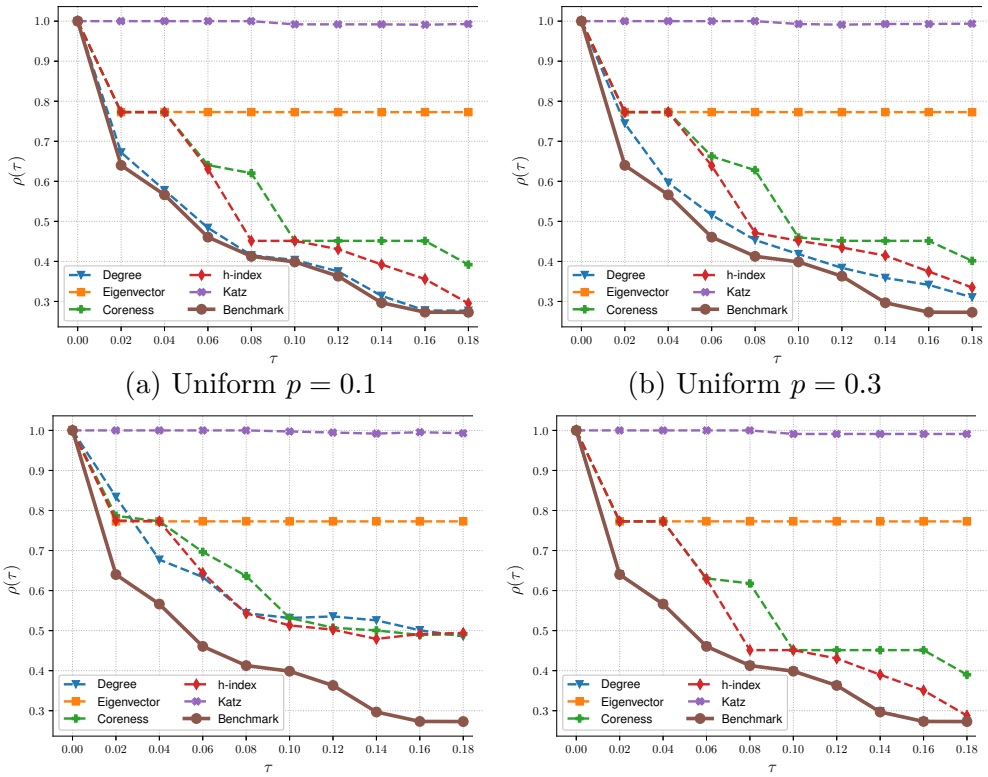
To verify the impact of our proposed models, λ_1 and c translate in two metrics, namely *effectiveness* and *coverage*, respectively. As extensively discussed in the previous sections, the first one is defined by $\rho(\tau, \phi)$ as the ratio of $\tilde{\lambda}_1$ to λ_1. Analogously, the *coverage* $\gamma(\tau, \phi)$ is defined as the ratio of $\tilde{c}$ to c. By construction, $\tilde{\lambda}_1 \leq \lambda_1$ and $\tilde{c} \leq c$, which implies that both $\rho(\tau, \phi)$ and $\gamma(\tau, \phi)$ always range between 0 and 1. The smaller the magnitude of $\rho(\tau, \phi)$ and $\gamma(\tau, \phi)$, the bigger the damage observed in the connectivity of G upon node removal. Those aspects will be commented in Sections 5.4.1 and 5.4.2, respectively.

Next, in Sect. 5.4.3, the effectiveness and coverage associated with the centrality metrics above have been compared with those of *NetShield* [55], a state-of-the-art approximation algorithm, which takes an integer k as input and seeks at discovering the set of k nodes which, if deleted from G, produce the biggest drop in λ_1.

Lastly, the key results obtained from our analysis will be summarised in Sect. 5.4.4.

5.4.1 Effectiveness of Centrality Measures

Figures 5.2-5.9 report the variation of effectiveness as a function of the fraction τ of removed nodes.

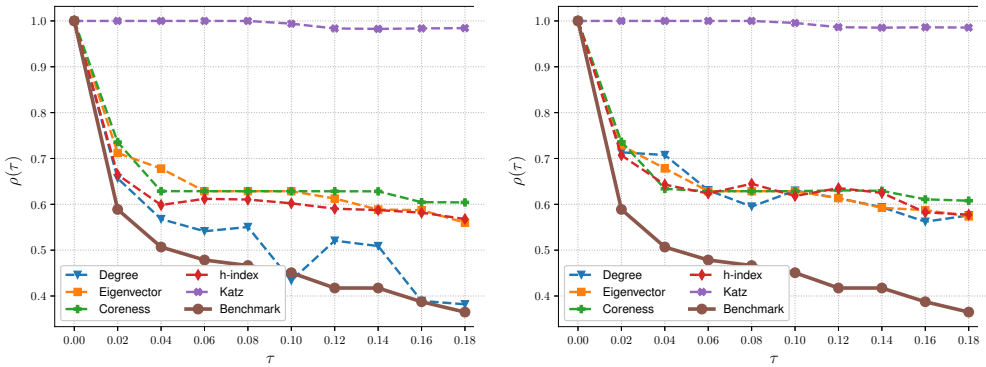


(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$

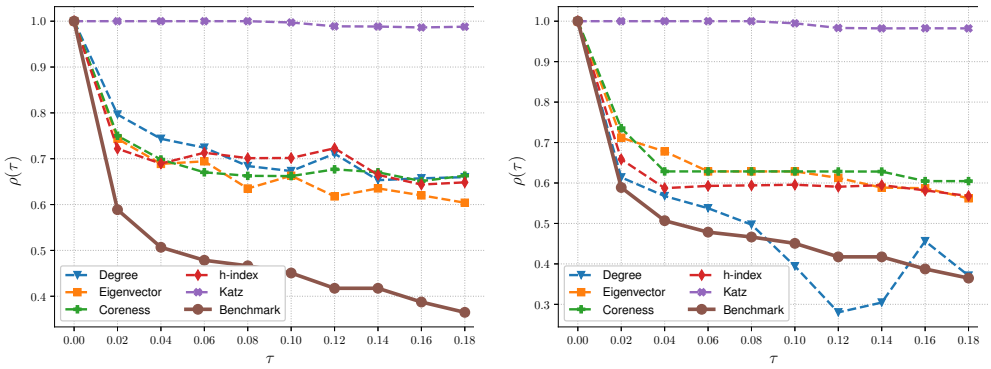
(c) Uniform $p = 0.5$ (d) BC

Figure 5.2: Effectiveness tests on FACEBOOK dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

In particular, Figure 5.2 shows the FACEBOOK dataset that, as will be later confirmed from the other experiments conducted on social networks, is quite fragile to node removal because of its topology. The Uniform model with $p = 0.1$, $p = 0.3$, and the BC one has been revealed to behave similarly. A slight lower drop is shown in $p = 0.5$, as expected. All of the scenarios above indicate that the degree is the most effective centrality metrics, followed by h-index and coreness. By contrast, the Katz centrality and Eigenvector have a negligible effect on network fragmentation. This is due to the network under scrutiny. Indeed, as explained in [161], this is an ego-network showing a typical behaviour where: some circles are contained completely within another one, others overlap with to each other, and a small portion of the remaining ones have no members in common with any other circle. Thus, metrics based on walks (such as Eigenvector and Katz) have a relatively small impact on this kind of network because most of the connections have short walks (or even cannot to go beyond their own circle).



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$



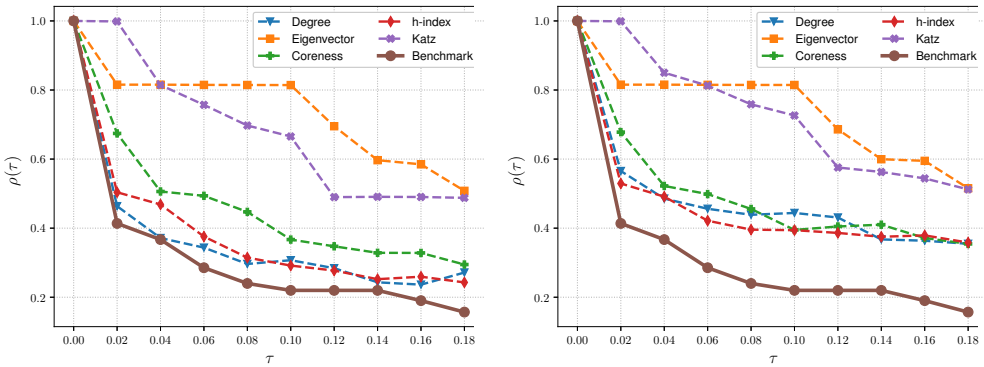
(c) Uniform $p = 0.5$ (d) BC

Figure 5.3: Effectiveness tests on US_Power_Grid dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

Figure 5.3 shows the variation of effectiveness in the US_Power_Grid dataset as function of τ. Considering the Uniform model with $p = 0.1$ and in the BC model, the degree gives the best effectiveness drop among all centrality metrics. In fact, it is sufficient to target a fraction $\tau = 0.02$ of nodes to lower the effectiveness from 1 to 0.63. If $p = 0.3$, then the degree achieve a comparable effectiveness with all the other metrics, apart from Katz. Similar behaviour has been encountered when $p = 0.5$. With Katz centrality, the reduction in effectiveness is almost negligible (around 0.01), for both the Uniform and the BC model. This effect derives from the fact that most of the nodes in US_Power_Grid have degree less than three; thus, the removal of large-degree nodes has a devastating impact on effectiveness.

LastFM experiments, displayed in Figure 5.4, can be seen as a kind of social network; thus, it behaves similarly to Facebook.

The collaboration network of ca-HepPh, shown in Fig. 5.5, diverges from the other experiments reported so far. Indeed, some differences are spotted among the centrality metrics tested with $\tau \geq 0.06$. However, as in the other analysis herein conducted, a comparable behaviour unites the Uniform model with $p = 0.1$ and the BC one. Moreover, as usual, the Uniform model with $p = 0.5$ has been



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$

(c) Uniform $p = 0.5$ (d) BC

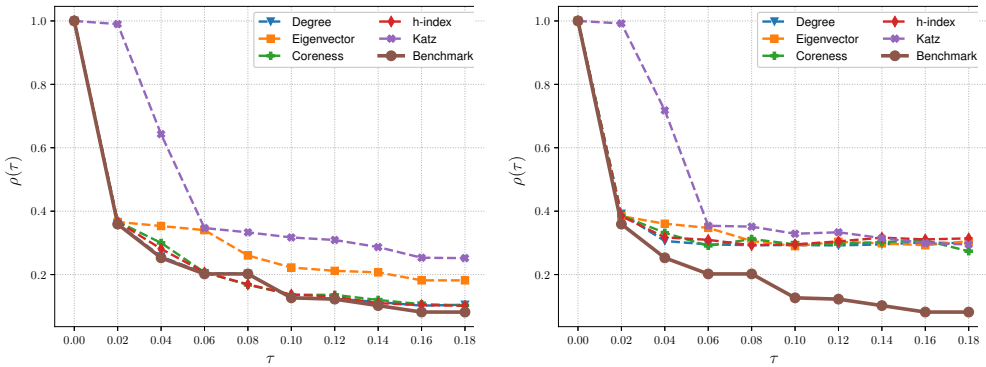
Figure 5.4: Effectiveness tests on LastFM dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

unveiled to be the most resilient. Indeed, when $\rho(\tau)$ is almost equal to 0.5 (*i.e.*, when $\tau \geq 0.06$) the curve becomes flatter.

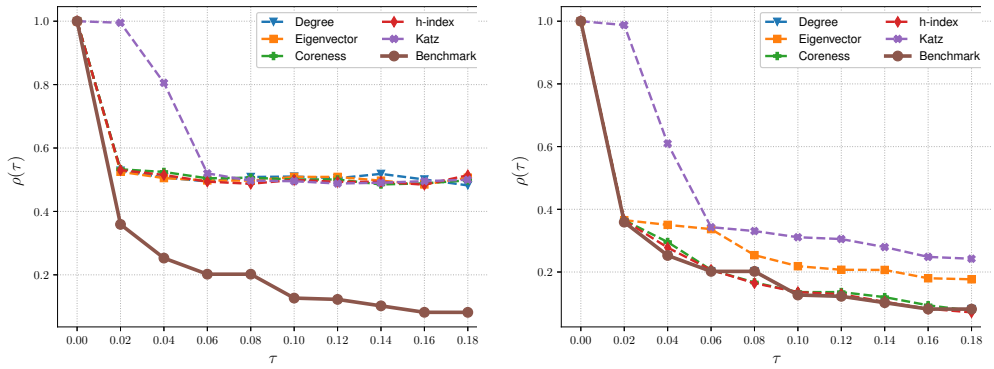
As for the AstroPh dataset (Figure 5.6), the degree is more effective than either Eigenvector and Katz centrality. Both h-index and coreness follow the same trend of the degree centrality. Furthermore, the decrease of effectiveness caused by the Katz centrality is smaller than that observed in case of the degree and in the Eigenvector centrality, and it stabilises for $\tau \geq 0.12$.

Next, we turn our attention to the Enron. In all experimental configurations (Fig. 5.7), the degree performs better than, or at least equal to, the other centrality metrics. If Katz centrality is applied, a slower decrease in effectiveness is observed. Indeed, as we previously found in the Facebook dataset, in the Uniform model with $p = 0.1$, it is sufficient to fix $\tau = 0.02$ to lower effectiveness to 0.24, which is about one third of the value measured in the case of the US_Power_Grid dataset.

In the BrightKite dataset the node removal effects are quite similar to the ones detected in the other social networks. Indeed, in Figure 5.8, Katz centrality performances prove to be worse than the others up to $\tau = 0.14$. For bigger



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$



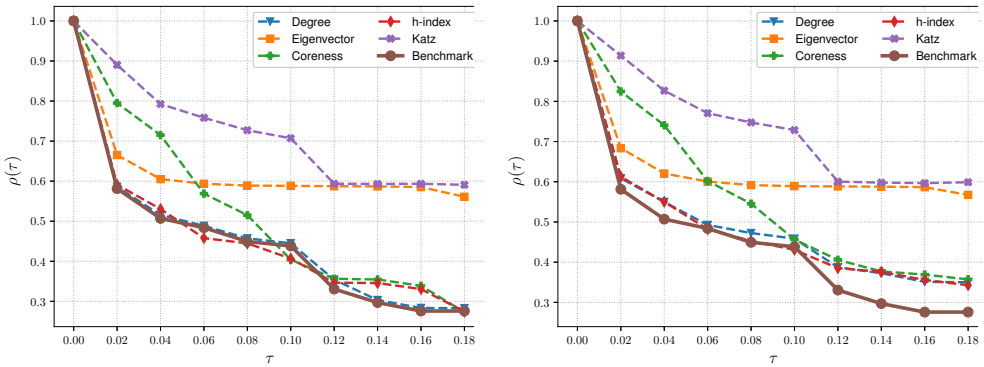
(c) Uniform $p = 0.5$ (d) BC

Figure 5.5: Effectiveness tests on CA-HEPPH dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

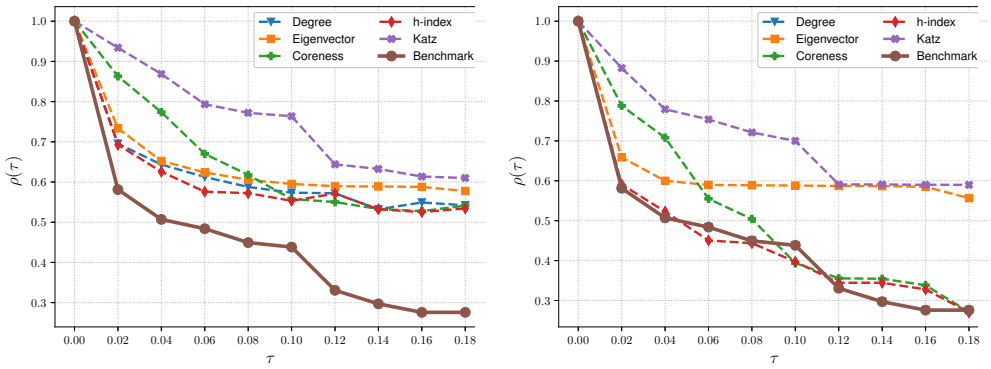
values of τ, modifications in the graph topology are so significant to cause a sharp decrease in effectiveness. Furthermore, the performances of the Uniform model with $p = 0.1$ have been also detected in BC model. In addition, the reduction in effectiveness in the Uniform model with $p = 0.3$ closely mirrors one with $p = 0.5$ with the clear exception that in $p = 0.3$ the drop is higher because of the higher network fragility.

FLICKR is the largest dataset herein investigated with more than two million edges. This may be regarded as a content network. However, since the process of producing metadata and associating them with images derives from the collaboration of Flickr users, it is predictable that FLICKR exhibits some features that make it similar to BRIGHTKITE. In fact, as shown in Figure 5.9, the degree generally yields the largest drop in the effectiveness.

Furthermore, effectiveness variations due to degree have been observed to be almost equal or, at least, they showed the same trend to the ones recorded through both h-index and coreness centrality, especially when graphs are sufficiently large. This was observed in all datasets herein investigated, both in Uniform and BC models. This result is consistent with the findings of Lu *et al.* [46], who high-



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$



(c) Uniform $p = 0.5$ (d) BC

Figure 5.6: Effectiveness tests on AstroPh dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

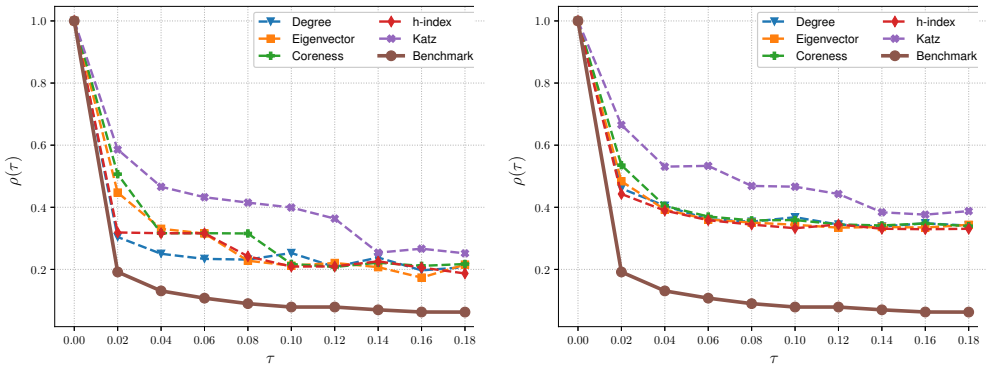
lighted a strong correlation between the degree, the h-index and the coreness.

An important exception arises with the US-Power-Grid graph. Indeed, most of the nodes exhibit an h-index of either one or two; thus, the process of choosing nodes on the basis of their h-index has a minor impact on the effectiveness.

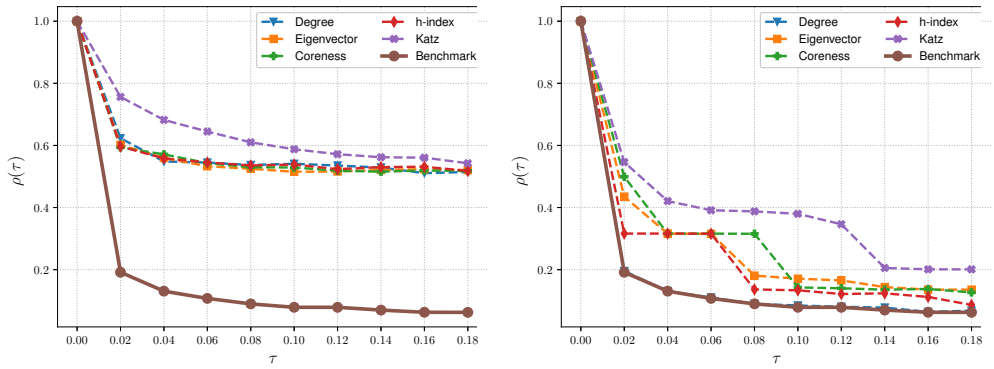
5.4.2 Coverage of Centrality Measures

Figures 5.10-5.17 report the variation of coverage as function of the fraction τ of removed nodes.

Figure 5.10 shows the coverage analysis conducted on the Facebook dataset. It is somehow complementary with the behaviour detected through the investigation of the effectiveness. In fact, the network topology justifies the big step drop spotted with $\tau \geq 0.10$ independently of the centrality metric herein used; indeed, it is likely that the network has been torn apart into isolated groups after the removal of nodes that acted as bridges among the circles. In the Uniform model with $p = 0.5$ this drop cannot be detected because the network resilience did not allow to remove those relevant nodes; thus, the trend is linear.



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$



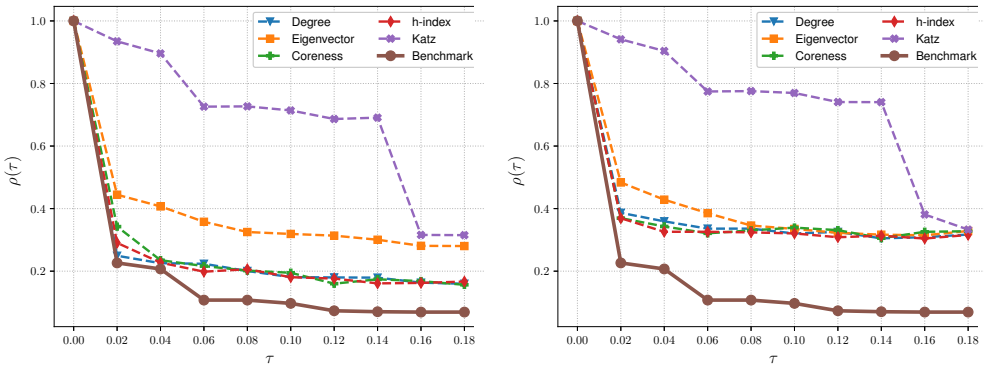
(c) Uniform $p = 0.5$ (d) BC

Figure 5.7: Effectiveness tests on ENRON dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

In the US_POWER_GRID dataset (Figure 5.11), the degree always yields the largest reduction in coverage. However, a similar trend is identified in both h-index and coreness even though the drop is slower than the degree centrality. Herein, we need to target a relatively large fraction of nodes (*i.e.*, $\tau \geq 0.14$) before observing a sensible reduction of coverage. Such a trend is likely dependent on the topological structure of power grid networks [5]. These kind of networks, in fact, display a high redundancy level; thus, they may endure the failure of a relatively small number of nodes before becoming disconnected.

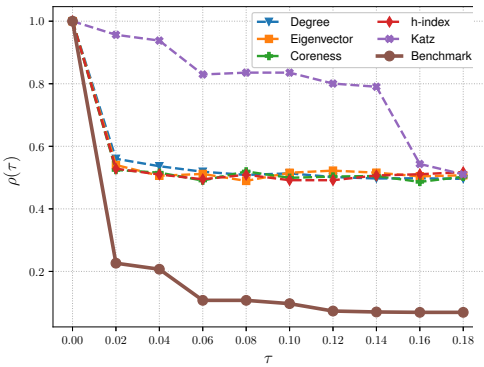
In the LASTFM, CA-HEPPH, ASTROPH, BRIGHTKITE and FLICKR datasets (Figures 5.12, 5.13, 5.14, 5.16, and 5.17, respectively), the coverage decreases in an almost-linear fashion, independently of the model adopted to encode node-failure probability and the centrality metric selected to target nodes. The Enron dataset (Fig. 5.15), even though with a flatter trend the other datasets, behaves similarly to them.

Once again, the degree is always responsible for the largest reduction in coverage, although in some configurations (*e.g.*, the FLICKR dataset in the Uniform model with $p = 0.5$) all centrality metrics cause the same amount of reduction in terms

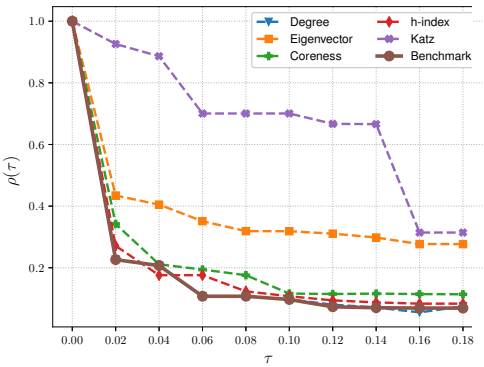


(a) Uniform $p = 0.1$

(b) Uniform $p = 0.3$



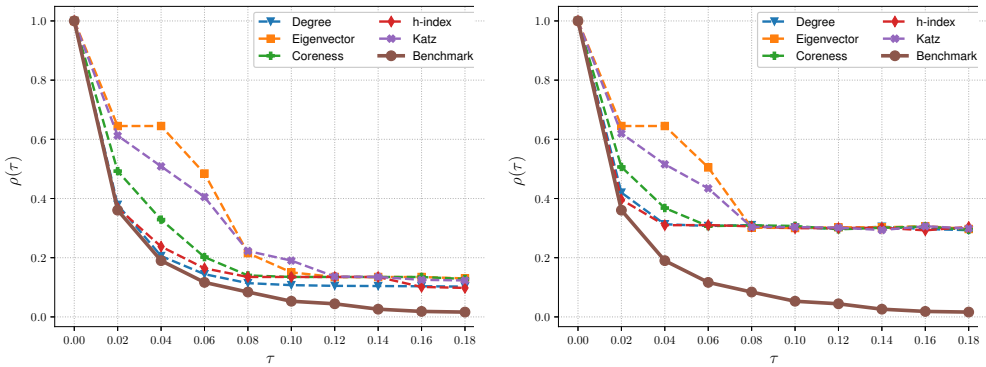
(c) Uniform $p = 0.5$



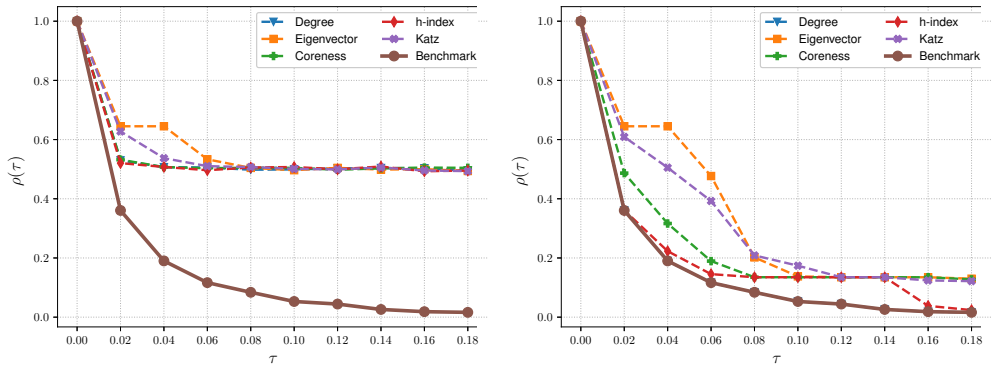
(d) BC

Figure 5.8: Effectiveness tests on BRIGHTKITE dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

of coverage. In particular, CA-HEPPH dataset (Fig. 5.13), the coverage dropped from 1 to 0.74 (in both the BC and Uniform model with $p = 0.1$). Analogously, the observed reduction in coverage in both ASTROPH and FLICKR (Figures 5.14 and 5.17) was from 1 to approximately 0.77 (both in the BC and Uniform models, with $p = 0.1$). LASTFM (Figure 5.12) shows the same behaviour with the only exception that its drop is slightly higher in both the BC and the Uniform model with $p = 0.1$. Furthermore, in the BRIGHTKITE dataset (Fig. 5.16) a stronger decrease in coverage emerged than in the LASTFM, CA-HEPPH, ASTROPH and FLICKR datasets. This is even more evident if the Uniform model with $p = 0.1$, or the BC model are considered. Be reminded that the BRIGHTKITE dataset edges identify friendship relationships among members. Therefore, there are few individuals who accumulate a large fraction of friendship relationships, and whose removal implies a quick fragmentation of the BRIGHTKITE graph into isolated components (similarly to what we noticed in the FACEBOOK ego-network).



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$



(c) Uniform $p = 0.5$ (d) BC

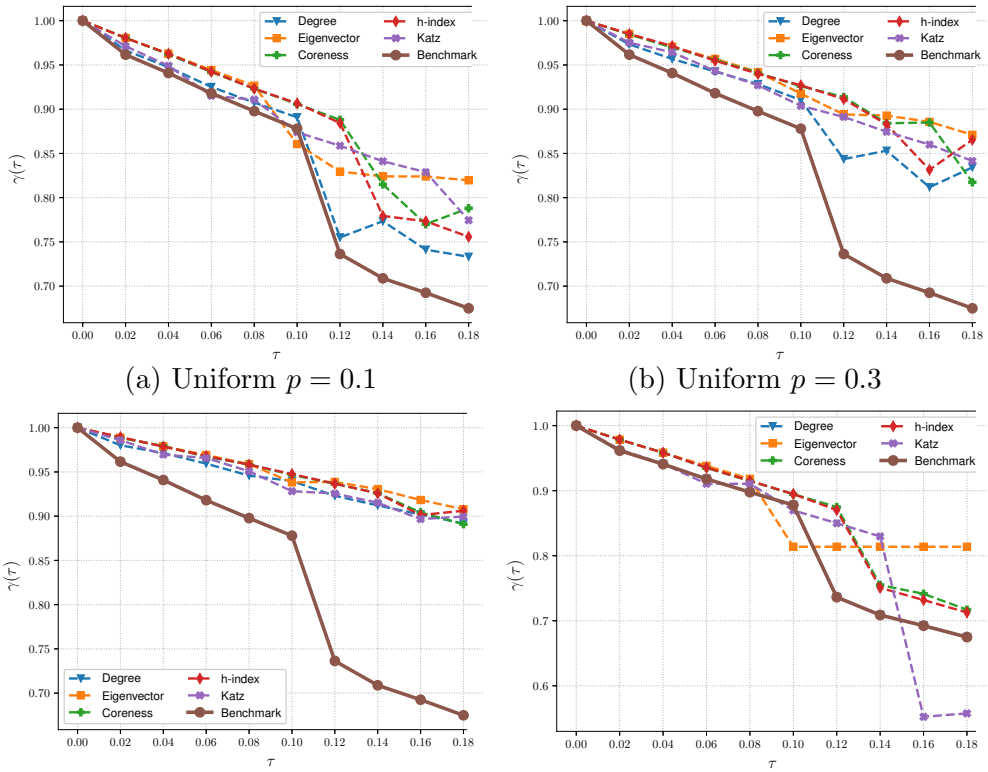
Figure 5.9: Effectiveness tests on FLICKR dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

5.4.3 A comparison with NetShield

To complete the analysis, this section compares the centrality metrics herein used and the NetShield algorithm [55].

As previously asserted, the NetShield algorithm takes an undirected graph G and an integer k as input and returns a set of k nodes $\mathcal{S}^{NS}(k)$ to be removed from G, aiming to achieve the biggest drop in λ_1.

To perform an experimental analysis, we had to slightly modify the evaluation protocol illustrated above. In fact, the worst-case time complexity of NetShield amounts to $O(nk^2+m)$, being n and m the number of nodes and edges of G: if, as in our previous tests, we assumed that k were of the same order of magnitude as n, then the worst-case time complexity of NetShield would have been cubic in n. This would make the application of NetShield unfeasible on even moderately large graphs. Therefore, we consider values of k ranging from 1 to 15. For the sake of brevity, we report our results only in the case of degree, which generally proved to be the most effective centrality metric among those we considered. To keep our notation consistent, we call $\mathcal{S}^d(k)$ the set of k nodes selected by the degree.



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$

(c) Uniform $p = 0.5$ (d) BC

Figure 5.10: Coverage tests on FACEBOOK dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

We introduce the parameter β to quantify the relative effectiveness of the NetShield algorithm against the degree:

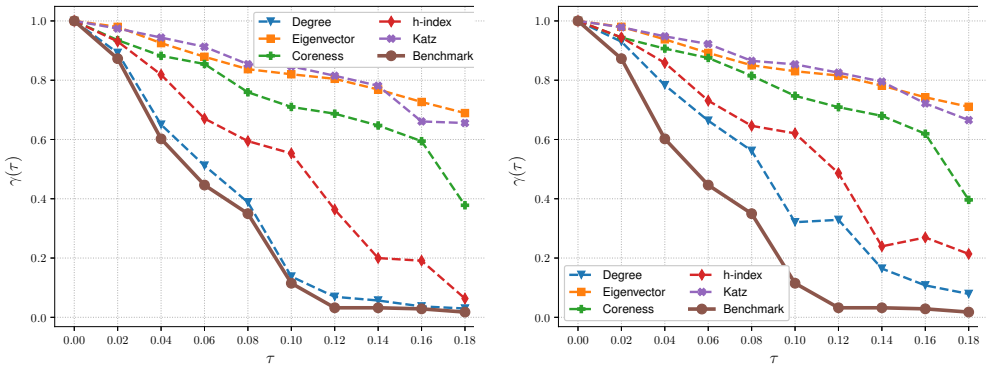
$$\beta = \frac{\tilde{\lambda}_1(NS)}{\tilde{\lambda}_1(d)} \tag{5.3}$$

Herein, $\tilde{\lambda}_1(NS)$ (resp., $\tilde{\lambda}_1(d)$) is the spectral radius of G after deleting nodes in $\mathcal{S}^{NS}(k)$ (resp., $\mathcal{S}^{d}(k)$).

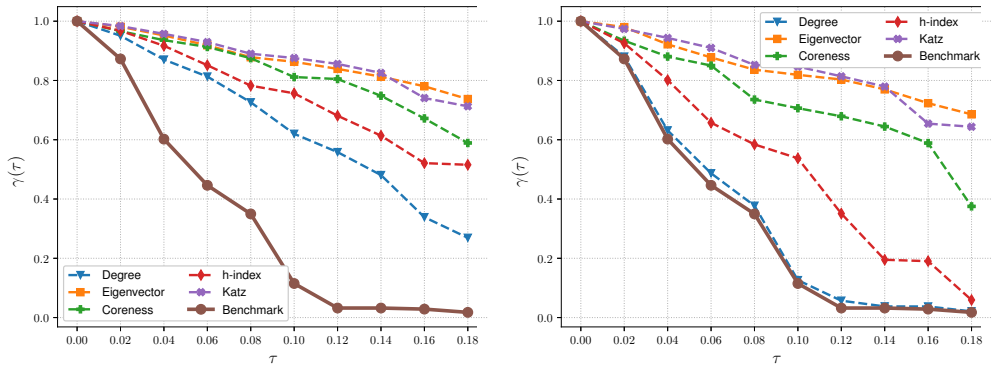
Analogously, we introduce the parameter ϵ to quantify the relative coverage of the NetShield algorithm against the degree:

$$\epsilon = \frac{\tilde{c}(NS)}{\tilde{c}(d)} \tag{5.4}$$

Herein, $\tilde{c}(NS)$ (resp., $\tilde{c}(d)$) is the LCC size of G after deleting nodes in $\mathcal{S}^{NS}(k)$ (resp., $\mathcal{S}^{d}(k)$). Observe that if $\beta < 1$ (resp. $\beta > 1$), then the NetShield algorithm is more (resp., less) effective than the degree. Analogously, if $\epsilon < 1$ (resp. $\epsilon > 1$), then the NetShield algorithm has a better (resp., worse) coverage (resp., less)



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$



(c) Uniform $p = 0.5$ (d) BC

Figure 5.11: Coverage tests on US_Power_Grid dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

effective than the degree.

The values of β in the Uniform and BC models (as k increases) are reported in Tables 5.2-5.5 (*i.e.*, $p = 0.1$ in Table 5.2, $p = 0.3$ in Table 5.3, $p = 0.5$ in Table 5.4, and BC in Table 5.5).

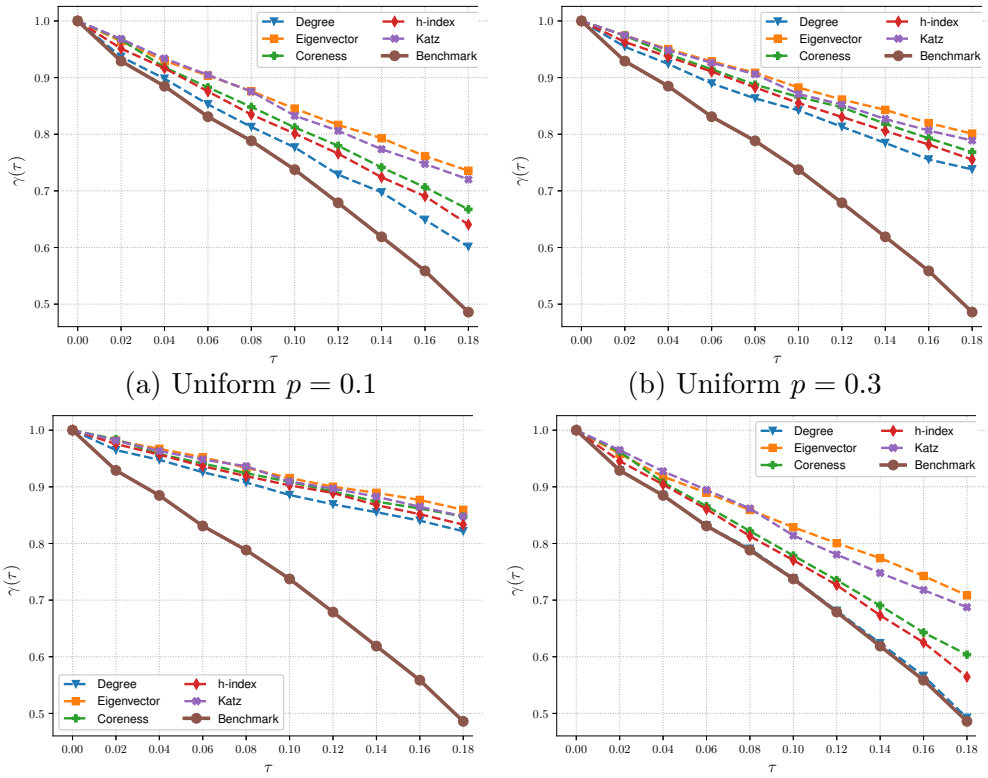
Table 5.2: Values of β as k increases in the Uniform model with $p = 0.1$.

k	Facebook	US_Power_Grid	LastFM	CA-Hep	AstroPh	Enron	BrightKite	Flickr
1	0.972	0.945	0.971	0.994	0.974	0.988	0.953	0.983
2	0.975	1.025	1.028	0.997	1.002	0.984	0.916	0.979
5	1.010	1.133	1.185	0.986	1.051	1.036	1.039	0.974
10	1.121	1.402	1.397	0.997	1.088	1.103	1.178	0.995
15	1.377	1.548	1.479	1.008	1.183	1.173	1.415	1.023

Table 5.3: Values of β as k increases in the Uniform model with $p = 0.3$.

k	Facebook	US_Power_Grid	LastFM	CA-Hep	AstroPh	Enron	BrightKite	Flickr
1	0.987	0.959	0.961	1.000	0.991	0.996	0.957	0.996
2	0.969	0.971	0.980	0.999	0.989	0.994	0.937	0.994
5	0.930	1.081	1.026	0.998	0.998	0.985	0.902	0.987
10	0.939	1.139	1.205	0.996	1.012	0.998	0.935	0.974
15	0.984	1.143	1.394	0.997	1.072	0.989	1.010	0.967

If we assume k less than or equal to five, then the β coefficient is always less than one and the lowest β values occur in the Uniform model with $p > 0.1$ as well as in



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$

(c) Uniform $p = 0.5$ (d) BC

Figure 5.12: Coverage tests on LastFM dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

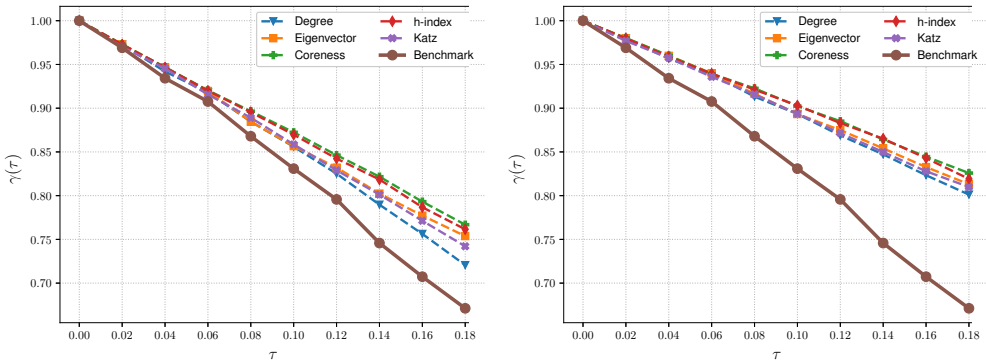
Table 5.4: Values of β as k increases in the Uniform model with $p = 0.5$.

k	Facebook	US_Power_Grid	LastFM	CA-Hep	AstroPh	Enron	BrightKite	Flickr
1	0.988	0.928	0.971	1.000	0.993	0.997	0.970	0.998
2	0.982	0.941	0.953	1.000	0.996	0.996	0.953	0.995
5	0.954	0.983	0.976	0.999	0.982	0.989	0.903	0.991
10	0.915	1.084	1.110	0.999	1.010	0.988	0.882	0.985
15	0.925	1.084	1.215	0.998	1.056	0.983	0.913	0.978

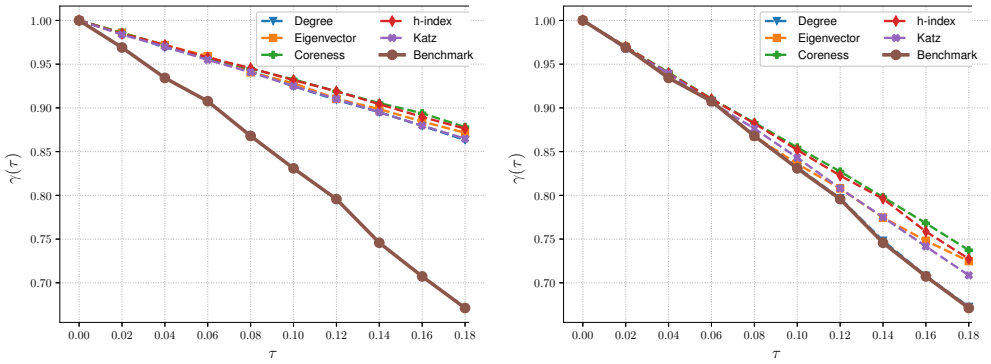
Table 5.5: Values of β as k increases in the BC model.

k	Facebook	US_Power_Grid	LastFM	CA-Hep	AstroPh	Enron	BrightKite	Flickr
1	0.992	0.926	1.000	1.0	0.998	1.000	0.979	0.999
2	0.992	0.919	0.952	1.0	1.000	0.998	0.984	0.999
5	0.981	0.979	0.906	1.0	0.997	0.997	0.966	1.000
10	0.959	0.995	1.041	1.0	1.010	0.997	0.969	0.996
15	0.945	1.056	1.129	1.0	1.020	0.996	0.933	0.994

the BC model. A significant exception occurs in two datasets (US_Power_Grid and Last-FM) where β is bigger than one. In concrete scenarios (*e.g.*, when we need to block the spread of an epidemics in a human community), the largest reduction in the spectral radius by blocking the least number of nodes is required. Thus, NetShield is the best weapon in our arsenal even if it can be time-consuming. The degree effectiveness is quite close to that of NetShield if $k < 10$, but the calculation of the degree is much faster than the application of the NetShield



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$



(c) Uniform $p = 0.5$ (d) BC

Figure 5.13: Coverage tests on CA-HEPPH dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

algorithm. Thus, the degree can be considered as a valid alternative to NetShield on large graphs.

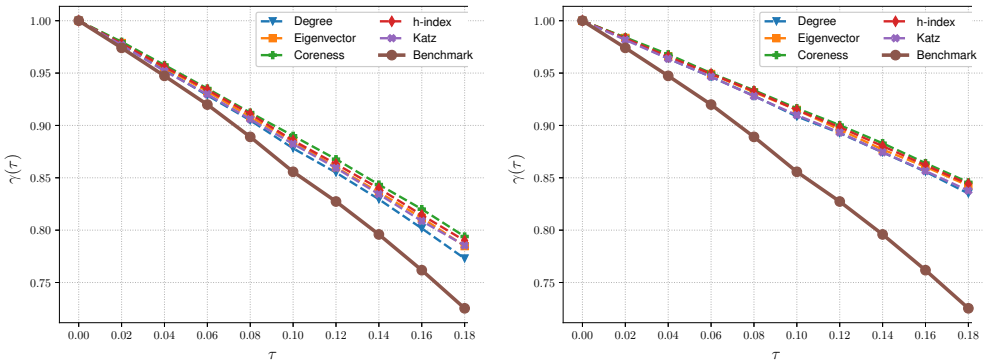
In contrast, if $k \geq 10$, the degree is more effective than NetShield, with an improvement up to 60%.

Let us now consider the relative coverage in Tables 5.6-5.9 (*i.e.*, $p = 0.1$ in Table 5.6, $p = 0.3$ in Table 5.7, $p = 0.5$ in Table 5.8, and BC in Table 5.9).

Table 5.6: Values of ϵ as k increases in the Uniform model with $p = 0.1$.

k	FACEBOOK	US_POWER_GRID	LASTFM	CA-HEP	ASTROPH	ENRON	BRIGHTKITE	FLICKR
1	1.026	0.957	1.014	1.001	0.997	1.006	1.170	1.000
2	1.034	1.070	1.105	1.005	1.022	1.006	1.166	1.000
5	1.121	1.350	1.380	0.995	1.019	1.042	1.334	1.000
10	1.286	1.556	1.810	1.015	1.047	1.018	1.232	1.001
15	1.099	1.800	2.912	1.097	1.042	1.078	1.592	1.001

In the Uniform model, the degree significantly outperforms NetShield on the US_POWER_GRID dataset, regardless of the p value. The ϵ value growth is proportional to k, which indicates the superiority of the degree in reducing the LCC size compared to NetShield. In the LASTFM dataset, we observe that the degree yields significantly larger values for ϵ than those achieved by the NetShield algo-



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$

(c) Uniform $p = 0.5$ (d) BC

Figure 5.14: Coverage tests on AstroPh dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

Table 5.7: Values of ϵ as k increases in the Uniform model with $p = 0.3$.

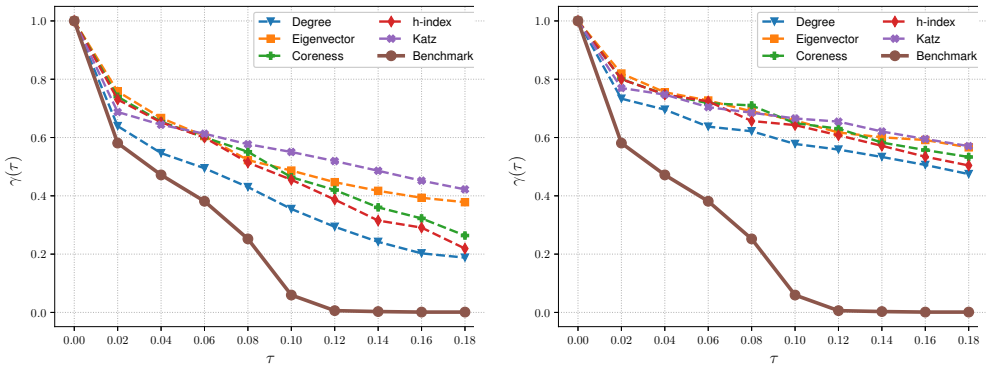
k	Facebook	US_Power_Grid	LastFM	CA-Hep	AstroPh	Enron	BrightKite	Flickr
1	1.011	1.021	1.009	1.000	1.000	1.001	1.007	1.000
2	1.019	1.056	1.010	1.000	1.001	1.002	1.009	1.007
5	1.034	1.109	1.024	1.001	1.000	1.005	1.015	1.000
10	1.077	1.375	1.033	1.002	1.004	1.006	1.021	1.006
15	1.051	1.542	1.072	1.002	1.003	1.009	1.026	1.000

Table 5.8: Values of ϵ as k increases in the Uniform model with $p = 0.5$.

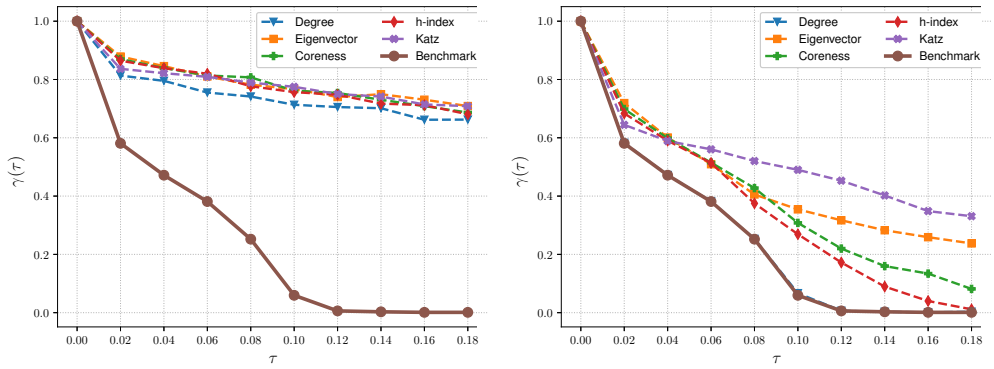
k	Facebook	US_Power_Grid	LastFM	CA-Hep	AstroPh	Enron	BrightKite	Flickr
1	1.002	1.001	1.000	1.0	1.000	1.001	1.003	1.000
2	1.030	1.038	1.003	1.0	1.000	1.001	1.004	1.005
5	1.009	1.384	1.006	1.0	0.999	1.006	1.009	1.000
10	1.048	1.717	1.012	1.0	1.001	1.006	1.010	1.001
15	1.017	1.583	1.015	1.0	1.000	1.005	1.012	1.000

Table 5.9: Values of ϵ as k increases in the BC model.

k	Facebook	US_Power_Grid	LastFM	CA-Hep	AstroPh	Enron	BrightKite	Flickr
1	1.236	1.012	1.0	1.0	1.0	1.0	1.0	1.0
2	1.237	1.012	1.0	1.0	1.0	1.0	1.0	1.0
5	1.237	1.012	1.0	1.0	1.0	1.0	1.0	1.0
10	1.239	1.014	1.0	1.0	1.0	1.0	1.0	1.0
15	1.240	1.020	1.0	1.0	1.0	1.0	1.0	1.0



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$



(c) Uniform $p = 0.5$ (d) BC

Figure 5.15: Coverage tests on ENRON dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

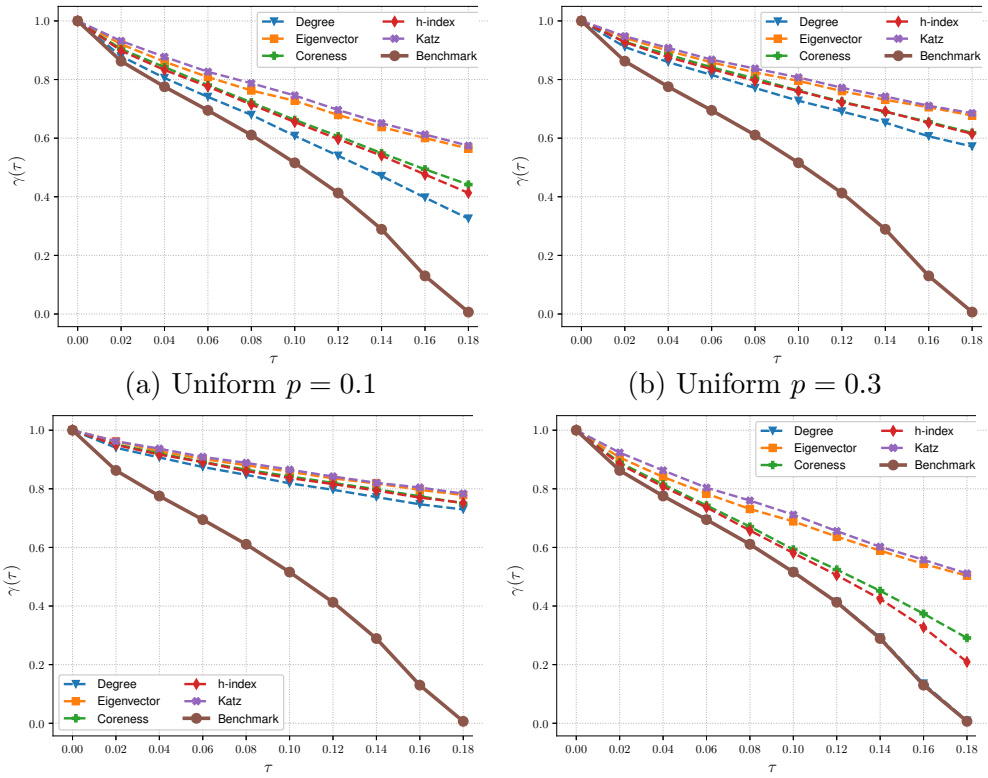
rithm if we opt for the Uniform model with $p = 0.1$; in contrast, the performance of the degree and NetShield are well aligned in the Uniform model with $p = 0.3$ and $p = 0.5$ and in the BC model.

The degree outperforms NetShield on the ASTROPH and BRIGHTKITE datasets in the Uniform model with $p = 0.1$. On the other hand, in the Uniform model with $p = 0.3$ and $p = 0.5$, ϵ values around one have been reported. Thus, the reduction in coverage due to the degree is almost equal to the one associated with Netshield.

In the BC model, experimental results highlight an almost perfect alignment between the coverage of NetShield and the one of the degree, confirmed by ϵ values close to one.

5.4.4 Summary of key results

In short, the take-home message from the experiments herein conducted is as follows: (i) Degree is a centrality metric that, on average, produces the largest drop in both λ_1 and c. Degree is, in addition, a viable alternative to other methods



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$

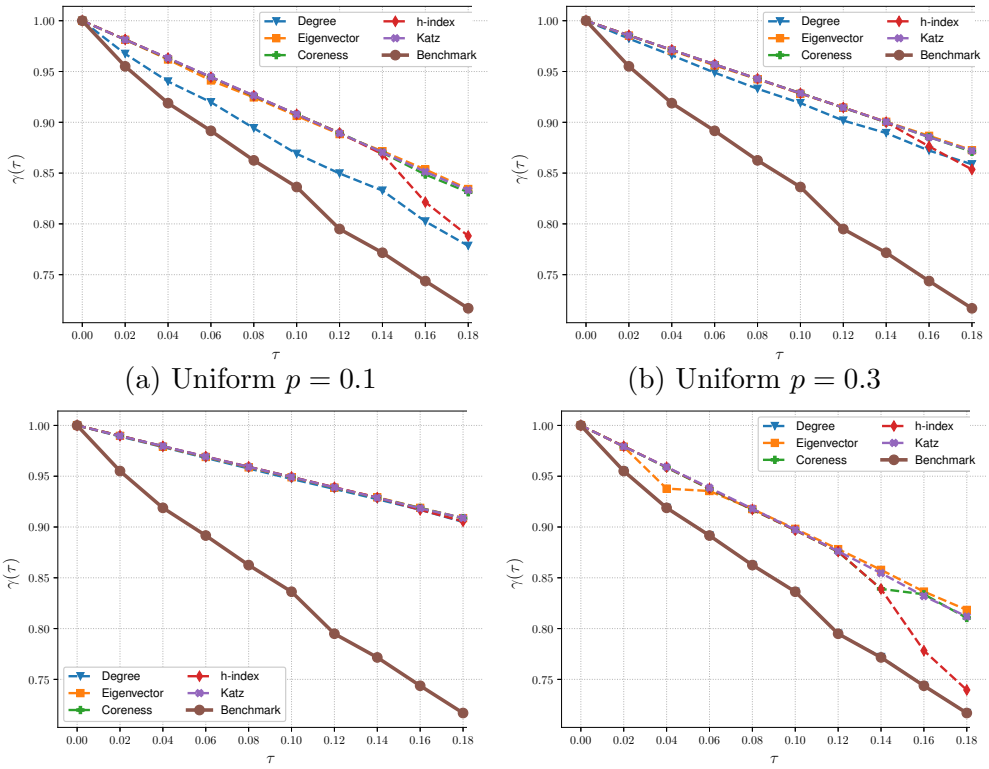
(c) Uniform $p = 0.5$ (d) BC

Figure 5.16: Coverage tests on BRIGHTKITE dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

(such as the NetShield algorithm) to detect group of nodes whose removal yields a relevant drop in the spectral radius. (ii) The BC model guarantees a large drop in λ_1, even when only a small fraction of nodes actually fail. (iii) In graphs deriving from human interactions and collaborations (namely FACEBOOK, LASTFM, CA-HEPPH, ASTROPH, ENRON, BRIGHTKITE, and FLICKR) a bigger drop in λ_1 rather than in US-POWER_GRID has been noticed. (iv) The degree, h-index, and coreness generally exhibit a similar trend in both social and collaborative networks, thus confirming insights provided in [46]. (v) The coverage analysis confirms that degree is better than the other centrality metrics herein considered, in terms of ability to fragment a graph into smaller and disjoint subcomponents, even when only a small fraction of nodes is targeted. In some datasets, a large gap in coverage reduction between degree and coreness centrality, and other centrality metrics has been appreciated; whereas in other datasets this gap appears to be more softened.

5.5 Discussion

This section illustrates the practical implications of our study.



(a) Uniform $p = 0.1$ (b) Uniform $p = 0.3$

(c) Uniform $p = 0.5$ (d) BC

Figure 5.17: Coverage tests on FLICKR dataset: (a-c) Uniform model with $p \in \{0.1, 0.3, 0.5\}$, (d) BC.

We observed that the survival probability p of a particular node in a graph G can be interpreted as the *cost* to remove the node. More specifically, the higher the survival probability of a node, the higher the cost of its removal.

In the Uniform model, the cost to remove a node is the same across all nodes. Therefore, this study suggests that *the choice of targeting high degree nodes is always the best one*, independently of the value of the survival probability p. If p were zero, the node removal task would always be successful. This case is well-known in the literature [149], and the conclusions of previous studies are consistent with our findings. However, in real–world systems (*e.g.*, the transportation system of a large city), high-degree nodes (often called *hubs*) are the most important points of failure; thus, these are typically adequately protected, so as to prevent cascading failures that would affect the whole system.

In the Best Connected (BC) model, the cost to remove a node is proportional to its degree. As an example, the cost to remove higher-degree nodes in power-law graphs can be some orders of magnitude bigger than the cost to remove lower-degree nodes. Suppose now to have a budget B to cover costs associated with the node-removal task, which is insufficient to remove as many nodes as desired.

Contrary to the Uniform model, the strategy of targeting high-degree nodes might not be optimal. Thus, a careful estimation of the trade-off between the costs of node removal and the corresponding loss in connectivity should be considered. For instance, Network Science methods have been widely applied to describe the structure of criminal organisations [60, 169, 126], and recent results indicate that high-degree nodes in a criminal organisation do not necessarily correspond to the major players in that organisation. Hence, a repressive action, which concentrates all the budget in removing high-degree nodes might even be ineffective, since the nodes corresponding to the major players have lower degree and would not be under attack. That is why in the case of criminal organisation an orderly node removal, starting from the higher ranks, is often ineffective. For the same reason, committing a whole budget, B toward the imprisonment of higher-degree nodes could just be a waste of time and financial resources.

Table 5.10 and Table 5.11 explicitly show the percent deviation between the Benchmark case ($p = 0$) and our probabilistic models, in terms of effectiveness and coverage, respectively. The reduction in the spectral radius λ_1 (resp., the size of the largest connected component c) of a graph observed in the Uniform/BC model vary broadly, depending on the type of dataset scrutinised, than that measured in a non-probabilistic model in which no node has the chance to survive an attack. However, the benefits tend to be substantial in most occasions. Indeed, it differs up to 82.44% from the classical approach.
Thus, these results show a significant difference between benchmark (conventional) and real-word situations (captured in our model).

Table 5.10: Difference between the probabilistic and classical approaches in computing effectiveness.

Dataset	$p = 0.1$	$p = 0.3$	$p = 0.5$	BC
Facebook	2.78%	10.9%	31.23%	0.07%
US_Power_Grid	2.02%	26.54%	35.82%	35.69%
LastFM	18.92%	41.9%	56.05%	0.72%
cah-HepPh	6.3%	46%	66.62%	0.03%
AstroPh	2.17%	11.39%	32.16%	22.42%
Enron	59.96%	74.4%	82.44%	3.71%
BrightKite	43.13%	66.57%	78.32%	62.26%
Flickr	44.07%	70.07%	80.31%	49.47%

Table 5.11: Difference between the probabilistic and classical approaches in computing coverage.

Dataset	$p = 0.1$	$p = 0.3$	$p = 0.5$	BC
Facebook	3.32%	8.42%	12.34%	0.001%
US_Power_Grid	24.82%	53.89%	65.88%	59.97%
LastFM	7.12%	14.71%	19.37%	0.49%
ca-HepPh	3.29%	7.71%	11.05%	0.1%
AstroPh	2.75%	6.16%	8.90%	4.28%
Enron	62.86%	71.12%	74.97%	9.23%
BrightKite	29.39%	39.34%	44.49%	28.43%
Flickr	4.88%	9.67%	12.31%	4.33%

5.6 Conclusions

In this chapter, a probabilistic model to describe node failure in graphs has been introduced, including two variants, dubbed Uniform and Best Connected (BC). Five popular centrality metrics have been considered (degree, h-index, coreness, Eigenvector, and Katz centrality), comparing their ability in reducing the spectral radius, λ_1 as well as the largest connected component size, c.

The main outcomes of our study are as follows: (i) the degree centrality metric, on average, generates the biggest drop in both λ_1 and c. Thus, we have also compared our outcomes with the always successful nodes removal approach, which from now on will be referred as "Benchmark", by ranking the nodes accordingly with the same centrality metric. In addition, the degree centrality is a viable alternative to the NetShield algorithm to quickly discover groups of nodes whose removal yields a relevant drop in the spectral radius. (ii) The BC model reports a large drop in λ_1 even when only a small fraction of nodes actually fails. To achieve a comparable reduction in λ_1 in the Uniform model we must, on average, delete more than 30% of nodes. (iii) In graphs deriving from human interactions and collaborations (namely Facebook, LastFM, ca-HepPh, AstroPh, Enron, BrightKite, and Flickr) a bigger drop in λ_1 rather than in US-POWER_GRID has been noticed. (iv) The degree, h-index, and coreness generally exhibit a similar trend in both social (such as BrightKite) and collaborative networks (such as AstroPh), thus confirming insights provided in [46]. (v) The degree yields the fastest decrease in coverage even though, in some datasets, such a decrease is only marginally smaller than that observed when the other centrality metrics discussed in this chapter are applied.

The node survival probability may also be interpreted as the cost needed to remove a node. Thus, our findings offer an opportunity to understand which nodes have

to be targeted/protected to deactivate a system or to keep it alive.

Finally, the comparison of our probabilistic model with the state-of-the-art, identifies significant differences, which confirms the value of our approach in real–world networks.

A possible extension of this project could be to include in this analysis the variation of centrality metrics in connection to edge removal [170]. A challenge will be to find a suitable probabilistic model that would mirror the performance of the Uniform and the BC strategies described in this chapter. Indeed, in real–world, Online Social Networks (OSNs), such as Facebook or Twitter, node failures reflect the deactivation of users' accounts, which is often a non-desirable effect. Yet, a range of other problems are linked to the interruptions in information flows, messages and status updates, which may be better captured by edge (rather than node) failures.

In addition, another extension of this project could be the use of criminal networks in order to compare the BC model considered herein with a new one that can be, instead, built to make the selection of targeted nodes proportional to their betweenness centrality score, instead of their degree (and their nodes resilience, accordingly). Thus, the challenge would be to define a proper benchmark to allow an appropriate comparison between the classical *always successful* approach and the probabilistic one herein described.

The next chapter will be the last experimental one. It will still consider the Network Features, but these will be explored under a rather different prospective. We will use the Network Science tools to investigate the role that the connections density could play in the context of Artificial Neural Networks (ANNs). Thus, starting from the Sparse Evolutionary Training (SET) framework, we went into analysing it with the goal to optimise the execution time, keeping a negligible loss in the accuracy of the training phase of an ANN, in the specific configuration of a Multi Layer Perceptron (MLP) topology.

Chapter 6

Artificial Neural Network Analysis

The work included in this chapter has been published in the following papers: [20, 21].

6.1 Introduction

The effort to simulate the human brain behaviour is one of the top scientific trends today. In particular, Deep Learning strategies pave the way to many new applications, thanks to their ability to manage complex architectures. Notable examples are: speech recognition [171], cyber-security [172], image [173] and signal processing [174]. Other applications gaining popularity are related to bio-medicine [175] and drug discovery [176, 177].

However, despite their success, Deep Learning architectures suffer from important scalability issues, *i.e.*, the actual Artificial Neural Networks (ANN) become unmanageable as the number of features increases.

While most current strategies focus on using more powerful hardware, the approach herein described employs Network Science strategies to tackle the complexity of ANNs iteratively, that is at each epoch of the training process.

This chapter originates from the work of Mocanu [15], a promising research avenue to speed up Neural Network training. There, a new approach called Sparse Evolutionary Training (SET) was defined, in which the acceleration effects obtained by

enforcing, in turn, scale-freeness, small-worldness, and sparsity, during the ANN training process, were explored.

Encouraged by those results, this chapter focuses on our research about looking at algorithm tuning parameters to pursue a further acceleration effect, at a negligible accuracy loss. The focus is on the revision stage (determined by the ζ parameter) and on its impact on the training time over epochs. Noteworthy results have been achieved by conducting an in-depth investigation into the optimal tuning of ζ, and by providing general guidelines on how achieve better trade-offs between time and accuracy, as described in Section 6.4.2.

6.2 Related Works

In recent years ANNs have been widely applied in a broad range of domains such as image classification [178], machine translation [179], and text-to-speech [180].

In [181] it was proved that the accuracy of an ANN (also known as *model quality*) crucially depends on both the model size (defined as the number of layers and neurons per layers) as well as the amount of training data. Due to these reasons, the amount of resources required to train large ANNs is often prohibitive for real-life applications.

A promising approach to achieve high accuracy even with modest hardware resources is *sparsity* [182]. An ANN is referred to as sparse when only a subset (hopefully of small size) of the model parameters has a value different from zero. The advantages of sparse networks are obvious. On one hand, sparse data structures can be used to store matrices associated with the representation of an ANN. On the other hand, most of the matrix multiplications (which constitute the most time-expensive stage of neural network computation) can be avoided. Furthermore, previous works [183, 15] suggested that high levels of sparsity do not severely affect the accuracy of an ANN.

6.3 Materials and Methods

This section firstly describes the datasets that were used to conduct our tests (Sect. 6.3.1). Then, it will be shortly illustrated the original SET framework from which the research herein presented originates (Sect. 6.3.2). Next, our innovative approach is described in Sect. 6.3.3. Lastly, in Sect. 6.3.4 our work will be compared with the original SET framework.

6.3.1 Dataset and ANN Descriptions

The experiments were conducted using well-known datasets, publicly available online[1]:

Lung Cancer: It is a biological dataset[2] composed by features on lung cancer in order to train the ANN to be able to detect them.

CLL_SUB_111: It is composed by B-cell chronic lymphocytic leukemia. This dataset[3] was born to profile the five most frequent genomic aberrations (*i.e.*, deletions affecting chromosome bands 13q14, 11q22-q23, 17p13 and 6q21, and gains of genomic material affecting chromosome band 12q13) [184].

COIL20: It is an image dataset[4] used to train ANNs to detect 20 different objects. The images of each object were taken five degrees apart as the object is rotated on a turntable and each object has 72 images. The size of each image is 32×32 pixels, with 256 grey levels per pixel. Thus, each one is represented by a 1024-dimensional vector [185, PAMI], [186, VLDB].

Both Lung Cancer and CLL_SUB_111 are biological datasets, widely used for their importance in medicine. Whereas the COIL20 dataset is a popular images dataset. Further quantitative details have been provided in Table 6.1.

Table 6.1: Dataset structures description. From left: dataset Name; dataset Type; Number of Instances, Number of Input Features; Number of Output Classes.

Name	Type	Inst. (#)	In. Feat. (#)	Out. C. (#)
Lung Cancer	Biological	203	3,312	5
CLL_SUB_111	Biological	111	11,340	3
COIL20	Face Image	1440	1024	20

The ANN used is composed of three hidden layers with 3,000 neurons per layer. The activation functions used by default are ReLu for the hidden layers, and Sigmoid for the output (See Table 6.2).

6.3.2 The SET framework

The SET framework, firstly initialises an ANN as a sparse weighted Erdős-Rényi graph in which the graph density is fixed ($\epsilon = 20\%$, by default), and assigns

[1] http://featureselection.asu.edu/
[2] https://sites.google.com/site/feipingnie/file/
[3] https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2466
[4] http://www.cad.zju.edu.cn/home/dengcai/Data/MLData.html

Table 6.2: Artificial Neural Network description. It provides information about: the Loss Function, the Batch sizes, the Learning rate, the Momentum and the weight decay.

Loss Function	MSE
Batch Size (fitting)	2
Batch Size (prediction)	1
Learning Rate	0.01
Momentum	0.9
Weight Decay	0.0002

weights to edges based on a normal distribution with mean equal to zero.

Then, at the end of each training epoch, there is a the revision step in which a fraction ζ of null-edges (*i.e.*, links with weight equal to zero) is selected and replaced with non-zero weights iteratively. This is done with the twofold goal of reducing the loss on the training set and to keep the number of connections constant. We should note that the revision step is not only rewiring the links but also re-computing the actual weight of the new links. The efficiency of this approach has also been recently confirmed by independent researchers, who managed to train a million-node ANN on non-specialised laptops [68].

6.3.3 Design of Experiments

To speed-up the training process the investigation relates to the effects drawn by ζ variations during the evolutionary weight phase, at each epoch. The analysis involves a gradual ζ reduction with the goal to provide a guideline on how to find the best ζ values range, to trade-off between speed-up and accuracy loss on different application domains.

In [15], the default revise fraction was set to $\zeta = 0.3$, (i.e, 30% of the revised fraction of nodes) and no further investigations on the sensitivity to ζ was carried out. Unlike Mocanu's research, an in-depth analysis on the revised fraction is herein conducted to understand these effects, particularly how the revise step affects the training when ζ is substantially reduced. Furthermore, note that in this chapter, $\zeta \in [0, 1]$ and $\zeta \in [0\% - 100\%]$ notations are used interchangeably.

Some obvious considerations of this problem are that a shorter execution time and a certain percentage of accuracy loss for smaller values of ζ are expected. Nonetheless, this relationship is bound to be non-linear; thus, it is crucial to get to quantitative results.

6.3.4 Comparison with the classical SET framework

In [15], the goal was to implement the SET algorithm and test it with numerous datasets, on several ANN types (MLPs, CNN, RBMs), and on different types of tasks (supervised and unsupervised learning). The current study investigates the role of the revise fraction parameter ζ, rather than on the algorithm itself. The aim is to provide a general guideline on finding the best ζ values range to reduce execution time, at a negligible loss of accuracy.

In [20], we suggested a preliminary study on the role of ζ having a negligible accuracy loss, lower fluctuations, and a valuable gain in overall execution time with $\zeta < 0.02$ with the Lung Cancer dataset. In the present chapter, this intuition is analysed on a wider range of datasets to provide stronger justifications for the findings. The most important contribution of our study has been to confirm the effectiveness of the SET framework. Indeed, the random sparseness in ANNs introduced by the SET algorithm is powerful enough even without further fine tuning of weights (*i.e.*, revise fraction) during the training process.

6.4 Results

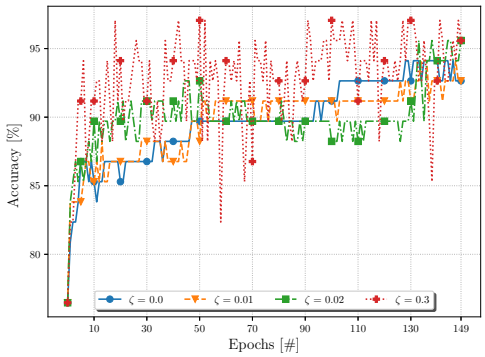
This section compares the results obtained by varying the parameter ζ, evaluating the training goodness in terms of the balance between high accuracy reached and short execution time. These topics are treated in Section 6.4.1 and Section 6.4.2, respectively. Section 6.4.3 provides a brief comment on the preferable ζ value, following up from the previous subsections.

For brevity, only the most important outcomes are reported hereafter. The number of epochs was increased from the default value of 100 up to 150 with the aim of finding the ending point of the transient phase. By combining these two tuning parameters (*i.e.*, number of epochs and ζ) we have discovered that, with the datasets herein analysed, the meaningful revise range is $0 \leq \zeta \leq 0.02$.

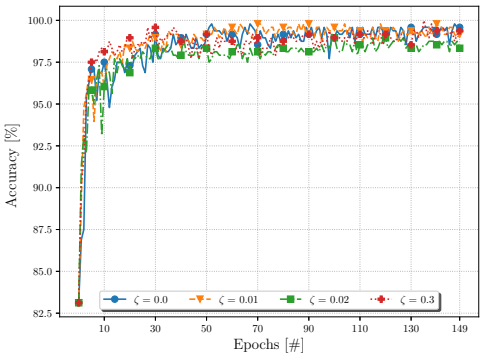
In particular, Section 6.4.2 shows further investigations in terms of execution-time gains, conducted by replicated experiments over ten runs and averaging the obtained results.

6.4.1 Accuracy Investigation

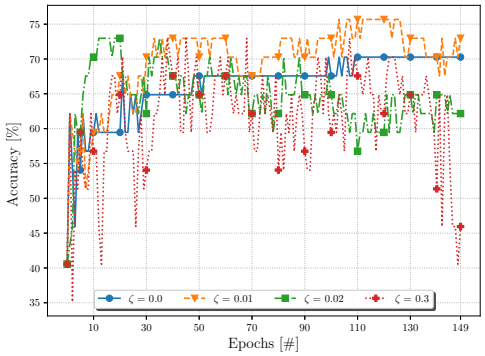
This section shows the results obtained from the comparative analysis in terms of accuracy improvements over 150 epochs, on the three datasets.



(a) *Lung Cancer* dataset



(b) *COIL20* dataset



(c) *CLL_SUB_111* dataset

Figure 6.1: Accuracy percentage over 150 epochs varying ζ among $[0\%, 1\%, 2\%]$ plus $\zeta = 30\%$ that is the benchmark value. In particular, $\zeta = 0\%$ with circled markers, $\zeta = 1\%$ has triangular markers, $\zeta = 2\%$ is shown with squared markers, and for $\zeta = 30\%$ cross shape markers have been used.

In the Lung Cancer dataset (Fig. 6.1a) substantial accuracy fluctuations are present, but there is a no well-defined transient phase for $\zeta > 0.02$. The benchmark value $\zeta = 0.3$, shows an accuracy variation of more than 10% (*e.g.*, accuracy increasing from 82% to 97% at the 60-th epoch, and an accuracy from 85% to 95% at the 140th epoch). Note that, since the first 10 epochs are within the settling phase, the significant observations concern the simulation from the 11th epoch. Due to this uncertainty, and due to the absence of a transient phase, it is impossible to identify an optimal stopping condition for the algorithm. For instance, at the 60th epoch an accuracy collapse from 97% to 82% was found, followed by an accuracy of 94% at the next epoch.

For a lower revise fraction, *i.e.*, $\zeta \leq 0.02$, an improvement in terms of both stability (*i.e.*, lower fluctuations) and accuracy loss emerges, as expected. In this scenario, defining an exit condition according to the accuracy trend over time is easier. Indeed, despite a higher accuracy loss, the curve stability allows the identification of a gradual accuracy growth over the epochs, with no unexpected sharp drops.

To quantify the amount of accuracy loss, refer to Table 6.3a, which reports both the revise fraction and the highest accuracy reached during the whole simulation, as a percentage. Moreover, mean and confidence interval bounds are provided. From Table 6.3a it is possible to assert that, on average, the improvement achieved by using a higher revise fraction (as the default one is) has an accuracy gain of just less than 3% (*e.g.*, mean at $\zeta = 0\%$ vs mean at $\zeta = 30\%$) that is a negligible improvement in most of the application domains. This depends on the tolerance level required. For example, if the goal is to achieve an accuracy of at least 90%, then a lower ζ is sufficiently effective. The confidence interval is rather low, given that the fluctuation between the lower and the upper bounds is comprised between 0.8 and 0.9.

Table 6.3: Evaluating parameters varying the revise fraction on datasets considered in a single run with fixed seed. From left: the revise fraction in percentage; the highest accuracy reached during the simulation expressed in percentage; the accuracy mean during the simulation, and the confidence interval bounds. Notes that these last three parameters are computed after the first 10 epochs to avoid noise.

(a) *Lung Cancer* dataset.

ζ (%)	Max Acc. (%)	Mean (%)	Lower B. (%)	Upper B. (%)
30%	97.06%	93.13%	92.67%	93.58%
2%	95.59%	90.38%	90.08%	90.68%
1%	94.12%	90.19%	89.84%	90.55%
0%	94.12%	90.21%	89.79%	90.62%

(b) *COIL20* dataset.

ζ (%)	Max Acc. (%)	Mean (%)	Lower B. (%)	Upper B. (%)
30%	100%	98.82%	98.75%	98.90%
2%	98.96%	98.17%	98.08%	98.25%
1%	99.79%	99.09%	98.99%	99.18%
0%	99.79%	98.84 %	98.70%	98.98%

(c) *CLL_SUB_111* dataset.

ζ (%)	Max Acc. (%)	Mean (%)	Lower B. (%)	Upper B. (%)
30%	72.97%	62.23%	61.14%	63.32%
2%	72.97%	65.15%	64.54%	65.76%
1%	75.67%	70.79%	70.15%	71.42%
0%	70.27%	67.14%	66.61%	67.67%

In the Coil20 dataset (Fig. 6.1b) a short transient phase with no evident improvements among the simulations with different values of ζ emerges. Indeed, there are just small accuracy fluctuations of $\pm 3\%$. These results do not surprise, since improvements achieved through ζ variations also depend on the goodness of the dataset itself, both in terms of its size and in the choice of its features.

Table 6.3b shows that accuracy is always above 98%; thus, even with $\zeta = 0$ the accuracy loss is negligible. Also the confidence interval is lower than 0.3. As the accuracy is continuously increasing over the training epochs, defining a dynamic exit condition is easier in this application domain.

Figure 6.1c shows the results obtained in CLL_SUB_111 dataset. It is evident that the worse and more unstable approaches among the one considered are both the default one (*i.e.*, $\zeta = 30\%$) and $\zeta = 2\%$.

From Table 6.3c it is interesting to notice how the accuracy levels are even more stable when using a lower revise fraction (*i.e.*, going from a mean equal to 62.23% in $\zeta = 30\%$ up to 67.14% in $\zeta = 0\%$). The fluctuations compared with the other two datasets are more evident, even when looking at the confidence interval; indeed, it varies from 1.06 (with $\zeta = 0$) up to 2.18 (with $\zeta = 30$), which is larger than the previously analysed one. Because of significant accuracy fluctuations, a possible early exit condition should be considered only with $\zeta = 0$ even at the cost of a slighter higher accuracy loss.

The results obtained so far suggest that there is no need to fine-tune ζ, because the sparsity introduced by the SET algorithm is sufficiently powerful, and only a few links need to be rewired (*i.e.*, $\zeta \leq 0.2$). Apart from the goodness of the datasets themselves (as in COIL20), opting for a lower revise fraction has shown that, from one hand, the accuracy loss in sometimes negligible. On the other hand, as it was in the CLL_SUB_111 dataset, the performances are even higher than the ones obtained though the benchmark value. This confirms the hypothesis made in Sect. 6.4.1 of the goodness of using a randomly sparse ANN topology.

6.4.2 Execution time Investigation

This section shows the comparative analysis conducted among the datasets used, in terms of execution time, over replicated simulations. Ten runs have been averaged, using the default value $\zeta = 0.3$, as benchmark (*i.e.*, $\zeta_{default}$). Note that only the most significant and competitive ζ value has been considered (*i.e.*, $\zeta_0 = 0$). Figure 6.2 shows the execution time (in seconds) of the same averaged simulations computed on the three datasets.

In both LUNG and CLL_SUB_111 datasets, $\zeta = 0$ is faster than the benchmark value. In particular, in CLL_SUB_111, the execution time is almost 40% faster than the default one and with higher accuracy performances too, as previously asserted in Section 6.4.1. It became less competitive in COIL20. The reason is the

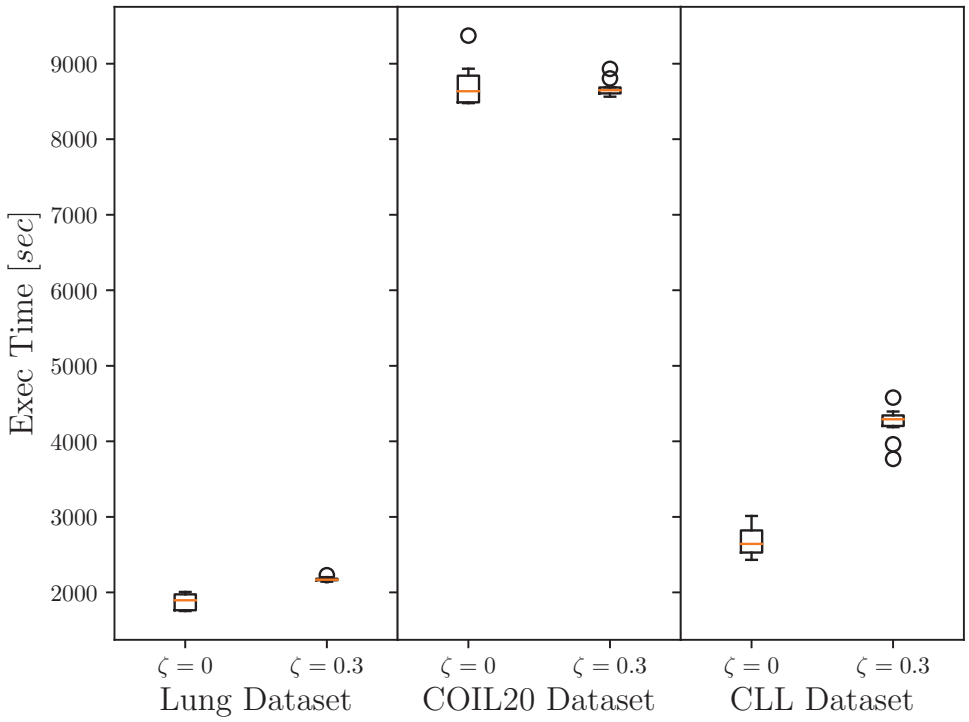


Figure 6.2: Execution Time over 10 runs. From left to right the Lung, COIL20 and CLL datasets are shown.

same with the results emerged in the accuracy analysis. Indeed, the goodness of the dataset is such as to make insignificant the improvements obtained by varying the revise parameter. Furthermore, the execution time gain between $\zeta = 0$ and $\zeta_{default}$ has been computed among the datasets over ten runs as follows:

$$Gain = 1 - \frac{\zeta_0}{\zeta_{default}} \tag{6.1}$$

The execution time gain was equal to 0.1370 in Lung, -0.0052 in COIL20, and 0.3664 in CLL_SUB_111. This means that, except for COIL20, there is an improvement in terms of algorithm performances. Thus, the algorithm became faster using a lower revise fraction. This is even more evident in CLL_SUB_111 as already noticed from Figure 6.2. On the other hand, the slow down emerged in COIL20 is almost negligible; thus, it may be concluded that for specific types of datasets, there is neither gain nor loss in choosing a lower ζ.

These results confirmed the previous hypothesis of the unnecessary fine-tune ζ process even because, on particular datasets (*e.g.*, COIL20), an in-depth analysis

of ζ is profitless. Thus, a relatively low revise fraction has been demonstrated to be a good practice in most of the cases.

6.4.3 Considerations on the ζ tuning process

In Sect. 6.4.1 and Sect. 6.4.2 we have described the effects of ζ in terms of accuracy loss and execution time, respectively. This section provides a brief summary of what emerged from those experiments. As largely discussed in the literature, it is unrealistic to try and find a perfect value, which works well in all possible Deep Learning scenarios *a priori*. The same consideration should be made during the revise fraction tuning. This is why those tests are not aimed at finding the optimal value, which depends on too many variables (*e.g.*, the input data, the configuration of the ANN in terms of number of hidden layers and number of neurons per each hidden layer, etc.). Instead, it may be asserted that, from the experiments herein conducted, a relatively low ζ seems to be a good choice.

In the analysed datasets, the best results have been obtained with $0 \leq \zeta \leq 0.02$. Our approach, thus, improves the ANN training because, when ζ is selected within the range we indicated, significant benefits are obtained. Indeed, the tests have been conducted on very different datasets to assert that, empirically speaking, the proper tuning of ζ is sufficient to offer a high accuracy with low fluctuations and, at the same time, faster execution time. It also important to highlight that because of the high non-linearity of the problem itself, more than one ζ value could effectively work, and the process of fine-tuning ζ is an operation that may require more time than the training process itself. This is why this study would provide a good enough range of possible ζ values (rather than a specific one): to provide a guideline for anyone interested in properly configuring a Sparse Artificial Neural Network.

6.5 Conclusions

In this chapter, we moved a step forward from earlier work [15]. Not only did our experiments confirm the efficiency arising from training sparse neural networks, but they also managed to further exploit sparsity through a better tuned algorithm, featuring increased speed at a negligible accuracy loss.

The revised fraction goodness is independent from the application domain, thus a relatively low ζ seems to be always a good practice. Of course, accordingly to the specific scenario considered, the performance may be higher than (or at least equal

to) the benchmark value. Yet, it is evident that Network Science algorithms, by keeping sparsity in ANNs, are a promising direction for accelerating their training processes.

From one side, acting on the revise parameter ζ, accuracy and execution time performances are positively affected. From the other side, it is unrealistic to try and define *a priori* an optimal ζ value, without considering the specific application domain, because of the high non-linearity of the problem. However, through this analysis it is possible to assert that a relatively low ζ is generally sufficient to balance both accuracy loss and execution time. Another strategy could be to sample the dataset in order to manage a lower amount of data and train only that portion of information on which to conduct tests on ζ.

This study paves the way for other works, such as the implementation of dynamic exit conditions to further speed-up the algorithm itself, the development of adaptive algorithms that dynamically tune the parameters, and the study of different distributions for the initial weight assignments.

Chapter 7

Discussion and Conclusions

This is the last chapter of the thesis. It is a comment of the main results obtained throughout the research herein presented (Sect. 7.1) jointly with a brief discussion of possible developments of this work (Sect. 7.2). Note that the source code on the criminal networks analyses herein presented is publicly available on a GitHub repository[1].

7.1 Summary of key results

The aim of this thesis was to show some of the potential of Network Science by detecting some of the most relevant features to be applied in different application domains, in particular in criminal networks. The focus was mainly on two topics: Criminal Networks and Artificial Neural Networks (ANNs), apparently unrelated to each other. This piece of research addressed four of the most significant problems in those areas to find, for the future, a connection between them.

In Chapt. 3 we investigated one of the most important problems that Law Enforcement Agencies (LEAs) have to deal with: arresting the lowest number of individuals to cause the highest damage to a criminal network. In particular, this thesis focused on the detection and application of network features to overcame this problem.

The first challenge to conduct this piece of research was related with data availability. Indeed, the difficulty was to obtain reliable data on which to conduct the

[1] https://github.com/lcucav/criminal-nets

simulations as these are typically sensitive data and, thus, LEAs cannot share such contents during their investigations.

Hence, we extracted the necessary information directly from juridical acts of concluded investigations and, then, we anonymised the sensitive information about criminal names to build two real–world datasets [125] that were related to two different perspectives of the same criminal organisation. In one case, telephone interceptions were considered (*i.e.*, Phone Calls dataset). In the other, the connections represented meetings between criminals (*i.e.*, Meeting dataset).

From those datasets, two weighted graphs, needed for our analyses, were obtained. Note that the networks relate to a "cosca" (*i.e.*, a clan or Sicilian Mafia crime family led by bosses) that operated in Sicily (Italy) during the first decade of the 2000s [187]. Next, we conducted a preliminary study on the networks topologies to have more information about criminal connections. In order to do so, we computed the weight distribution and the shortest path length analyses, from which it emerged that there were only a few criminals having a higher interaction frequency, and that there were a balanced number of intermediates to spread a message within the networks under scrutiny.

The comparative analysis between the two networks unveiled that there was a lower average interaction frequency in phone communications. This showed that criminals tend to minimise their interactions to keep the interception risks low, and that phone calls overexpose them, as they cannot know whether or not their devices are under surveillance. For the same reason, the well-connected criminals were not the bosses, even though they play a key role in connecting criminals.

With this knowledge in mind, the problem of detecting the most relevant suspects (*i.e*, the ones that, if arrested, produce a significant slowdown in information spreading within the criminal organisations) was addressed. We ranked the suspects using different centrality metrics (*i.e.*, Degree, Betweenness, Katz, Collective Influence) and we removed them from the networks, accordingly.

To evaluate the network disruption effect, the Largest Connected Component (LCC) size drop was computed, both before and after the node removal process. Those analyses unveiled the Betweenness centrality as the most effective metric (*i.e.*, simulating the arrest of only 5% of the top-ranking affiliates, the LCC dropped by 70%.) rather than by individual node Degree, thanks to its prioritisation of communication paths. This procedure tackled directly the communication mechanisms used by criminal networks, which are generally designed to minimise

the probability of interception by the police. This piece of research led to two papers [16, 17].

Another hard problem relates dealing with criminal networks affected by missing (or inconsistent) data. Thus, in Chapt. 4 we addressed the missing data problem on criminal networks. A simulation on the lack of information was performed by a comparative analysis between nine real criminal networks (summarised in Table 4.3), which were obtained from six investigation datasets, and their pruned counterparts. Two types of missed information were considered: (i) unavailability of sporadic interactions among suspects (*i.e.*, random edges removal from the graphs); (ii) impossibility to investigate specific suspects, for instance, when there is a lack of proof (*i.e.*, nodes removal from the graphs).

To quantify and evaluate the threshold of network comprehension under those perspectives, four distance metrics were used (*i.e.*, Adjacency, Laplacian, normalised Laplacian and Rood Euclidean distances). We pruned up to 10% of the overall edges in the first case, or nodes (and their incident edges) in the second one.

The most significant outcome that emerged from our analyses was that the network similarity between the original network and the pruned one is still high, even when a quite large amount of interceptions were missed (*i.e.*, 10% of random removed edges led to 30% of difference from the original graphs, compared with the 80% drops in case of removed nodes). This investigation led to a journal paper [16].

One of the main limitations in this kind of studies is that Machine Learning techniques cannot be applied as they are nowadays, because a huge amount of reliable and complete criminal datasets would be needed to properly train an Artificial Neural Networks (ANNs).

In Chapt. 5, another issue when dealing with networks simulations was addressed. In particular, to recreate the characteristic of network resilience a new framework was theorised: the probabilistic failure model. Indeed, removing nodes from a graph is generally performed as an always successful process; however, in real–word scenarios, it could not be as straightforward. Note that the failure can be generated from both an internal or from an external event. In the first case, the cause may be, for instance, a malfunction (*e.g.* power-grids and consequent blackouts). In the second one, an intentional attack may occur. Herein, the graphs considered were both human engineered (that are resilient by construction) and social networks to allow a comparative analysis between those topologies.

Eight real–world graphs were considered in total and the model proposed had two

variations (i) *Uniform* (the survival-to-failure probability was the same for all the nodes in the network); and (ii) *Best Connected* or BC (the survival probability was proportional to node degree). The nodes were first ranked according to five of the most popular centrality metrics (*Degree*, *H-Index*, *Coreness*, *Eigenvector*, and *Katz*); afterwords the removal attempts were preformed. To evaluate the goodness of the approach, two metrics were considered: (i) *effectiveness* (*i.e.*, the drop in the spectral radius λ_1 after node removal) and (ii) *coverage* (*i.e.*, the reduction c of the LCC size of a graph).

The outcomes from those analyses unveiled significant differences from the BC version, the state-of-art, and our benchmark strategy (*i.e.*, always successful node removal through degree centrality). Our hypothesis that the non-probabilistic approach is often unrealistic was confirmed, because the benchmark did not consider the cost of the node removal process. In turn, this often leads to incorrectly forecasting the speed of network fragmentation. This piece of research has not been published yet. While it is undergoing peer-review in IEEE Transactions on Network Science and Engineering, it is accessible in ArXiv [19].

One of the main limitations of this piece of research is that this model cannot be applied on criminal networks as is. The challenge is the definition of a proper benchmark to allow an appropriate comparison between the classical *always successful* approach and the probabilistic failure one herein described.

Finally, Chapt. 6 was the last experimental one, which addressed the problem of optimising the training of Sparse Artificial Neural Networks. As discussed in this thesis, ANNs are generally used as a black-box in which some inputs are provided and some outputs are returned in turn. However, it is still not known how those networks work and why they are so accurate (when properly modelled). In addition, those networks perform better with resource-consuming configurations, which require High Performance Computing (HPC) power, and a huge amount of input information is needed.

To streamline those structures without (or, at least, with a negligible) accuracy loss, an analysis of the Sparse Evolutionary Training (SET) framework [15] was performed. In particular, the selection criterion of the rewiring fraction size was analysed on three publicly available datasets. The analyses unveiled that this step can be avoided (*i.e.*, no rewiring phase is needed in most of the cases) still keeping the accuracy loss negligible. This discovery sped up the execution time, and the computational cost was reduced consequently. These studies have been published in [20] and [21].

Unfortunately, as previously asserted, mainly due to the lack of data availability, it is almost impossible nowadays to apply such techniques on criminal networks as they are. However, the promising results of this preliminary study allow further researches in this direction.

7.2 Future directions

This thesis is prone to considerable extensions and adaptations. One of the most significant limitations relates to the insufficiency of the use of the centrality metrics on their own for analysing criminal networks.

Indeed, even though the Betweenness centrality was effective to detect key criminals that, if arrested, cause a significant slow–down in the information spreading, detecting the bosses through the existing metrics is still far from the reality. Thus, new metrics have to be considered to gain better insights into Mafia clans interconnections that, as asserted throughout this thesis, communicate differently from other social networks.

One possibility would be to combine popular centrality metrics with new ones that better capture these anomalous types of communications. For instance, a combination of social (network characteristics in terms of connections, detectable through Graph theory) and human (personal skills and competences of the members, invisible through Social Network Analyses) capitals may be considered. Under this prospective, the use of temporal networks may be crucial for detecting patterns by analysing the evolution of the organisation. For instance, a focus on the role that small traders (*e.g.* greengrocers or bakers) play in facilitating the interactions within and outside the Families mobsters could be a successful strategy to discover such communication patterns and, in turn, differentiate between (i) unrelated to the Mafia context (*e.g.* mobsters who occasionally buy something) and (ii) requests for protection (*i.e.* "pizzo" / racket) by the traders, which is typically periodical.

Another interesting aspect to investigate as future extension of this project is a wider investigation on the probabilistic failure model including edges removals [170]. In real–world Online Social Networks (OSNs), for instance, node failures reflect the deactivation of users' accounts, which is often a non-desirable effect. Yet, a range of other problems are linked to the interruptions in information flows, messages and status updates, which may be more efficiently captured by edge (rather than node) failures. Consequently, acquiring such knowledge on network resilience

properties may help also further developments onto the criminal networks investigations direction.

For the same reason, a comparative analysis between criminal and general social networks may be a fruitful investigation also in terms of pruned networks. This can allow a clearer understanding of the effects of missed information.

Finally, this thesis paved the way for projects that may, in the future, allow the use of Artificial Neural Networks tools on Digital Forensic domain. For instance, the implementation of dynamic exit conditions to further speed-up the algorithm itself, the development of adaptive algorithms that dynamically tune the parameters, and the study of different distributions for the initial weight assignments may be possible the definition of specific artificial networks able to accurately simulate and predict the behavioural patterns of real criminal organisations.

Bibliography

[1] C. Morselli, Inside Criminal Networks, Springer, New York, US, 2008. doi:10.1007/978-0-387-09526-4.

[2] M. Peterson, Applications in Criminal Analysis: A Sourcebook, Greenwood Press, 1994.

[3] D. Koutra, J. T. Vogelstein, C. Faloutsos, Deltacon: A principled massive-graph similarity function, Proc of the 2013 SIAM International Conference on Data Mining (2013). doi:10.1137/1.9781611972832.18.

[4] R. Albert, H. Jeong, A. L. Barabási, Error and attack tolerance of complex networks, Nature 406 (6794) (2000) 378–382. doi:10.1038/35019019.

[5] R. Alber, I. Albert, G. Nakarado, Structural vulnerability of the North American power grid, Physical review E 69 (2) (2004) 2, 025103. doi:10.1103/PhysRevE.69.025103.

[6] C. Chen, H. Tong, B. Prakash, T. Eliassi-Rad, M. Faloutsos, C. Faloutsos, Eigen-optimization on large graphs by edge manipulation, ACM Transactions on Knowledge Discovery from Data (TKDD) 10 (4) (2016) 49. doi:10.1145/2903148.

[7] Y. Wang, D. Chakrabarti, C. Wang, C. C. Faloutsos, Epidemic spreading in real networks: An eigenvalue viewpoint, in: Proc of the International Symposium on Reliable Distributed Systems (SRDS 2003), IEEE, 2003, pp. 25–34. doi:10.1109/RELDIS.2003.1238052.

[8] A. Ganeshi, L. Massoulié, D. Towsley, The effect of network topology on the spread of epidemics, in: Proc of the Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2005), IEEE, 2005, pp. 1455–1466. doi:10.1109/INFCOM.2005.1498374.

[9] B. Prakash, D. Chakrabarti, N. Valler, M. Faloutsos, C. Faloutsos, Threshold conditions for arbitrary cascade models on arbitrary networks, Knowledge and Information systems 33 (3) (2012) 549–575. doi:10.1007/s10115-012-0520-y.

[10] N. Berger, C. Borgs, J. Chayes, A. Saberi, On the spread of viruses on the internet, in: Proc of the ACM-SIAM Symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 2005, pp. 301–310.

[11] J. Kleinberg, The wireless epidemic, Nature 449 (7160) (2007) 287. doi:10.1038/449287a.

[12] J. Jiang, S. Wen, S. Yu, Y. Xiang, W. Zhou, Identifying propagation sources in networks: State-of-the-art and comparative studies, IEEE Communications Surveys & Tutorials 19 (1) (2017) 465–481. doi:10.1109/COMST.2016.2615098.

[13] M. Amoruso, D. Anello, V. Auletta, D. Ferraioli, Contrasting the spread of misinformation in online social networks, in: Proc of the 16th Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2017), 2017, pp. 1323–1331. doi:10.1613/jair.1.11509.

[14] D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, The science of fake news, Science 359 (6380) (2018) 1094–1096. doi:10.1126/science.aao2998.

[15] D. C. Mocanu, E. Mocanu, P. Stone, P. H. Nguyen, M. Gibescu, A. Liotta, Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science, Nature communications 9 (2018) 1–12. doi:10.1038/s41467-018-04316-3.

[16] A. Ficara, L. Cavallaro, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, A. Liotta, Social network analysis of Sicilian Mafia interconnections, in: H. Cherifi, S. Gaito, J. Mendes, E. Moro, L. Rocha (Eds.), Complex Networks and Their Applications VIII, Springer International Publishing, Cham, Switzerland, 2020, pp. 440–450. doi:10.1007/978-3-030-36683-4_36.

[17] L. Cavallaro, A. Ficara, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, W. Song, A. Liotta, Disrupting resilient criminal networks through data analysis: The case of Sicilian Mafia, PLoS ONE 15 (8) (2020) 1–22. doi:10.1371/journal.pone.0236476.

[18] A. Ficara, L. Cavallaro, F. Curreri, G. Fiumara, P. De Meo, O. Bagdasar, W. Song, A. Liotta, Criminal networks analysis in missing data scenarios through graph distances, PLoS ONE 16 (8) (2021) 1–18. doi:10.1371/journal.pone.0255067.

[19] L. Cavallaro, S. Costantini, P. De Meo, A. Liotta, G. Stilo, Network connectivity under a probabilistic node failure model (2020). arXiv:2006.13551.

[20] L. Cavallaro, O. Bagdasar, P. De Meo, G. Fiumara, A. Liotta, Artificial neural networks training acceleration through network science strategies, in: Y. D. Sergeyev, D. E. Kvasov (Eds.), Numerical Computations: Theory and Algorithms, Springer International Publishing, Cham, Switzerland, 2020, pp. 330–336. doi:10.1007/978-3-030-40616-5_27.

[21] L. Cavallaro, O. Bagdasar, P. De Meo, G. Fiumara, A. Liotta, Artificial neural networks training acceleration through network science strategies, Soft Computing 24 (2020) 17787–17795. doi:10.1007/s00500-020-05302-y.

[22] L. Linyuan, T. Zhou, Link prediction in complex networks: A survey, Physica A: Statistical Mechanics and its Applications 390 (6) (2011) 1150–1170. doi:10.1016/j.physa.2010.11.027.

[23] M. A. Hasan, M. J. Zaki, A survey of Link Prediction in Social Networks, in: C. Aggarwal (Ed.), Social Network Data Analytics, Springer US, Boston, MA, 2011, pp. 243–275. doi:10.1007/978-1-4419-8462-3_9.

[24] L. Tassiulas, D. Katsaros, P. Basaras, Detecting influential spreaders in complex, dynamic networks, Computer 46 (4) (2013) 24–29. doi:10.1109/MC.2013.75.

[25] U. Bellur, R. Kulkarni, Improved matchmaking algorithm for semantic web services based on bipartite graph matching, in: I. International (Ed.), IEEE International Conference on Web Services (ICWS 2007), Salt Lake City, UT, 2007, pp. 86—93. doi:10.1109/ICWS.2007.105.

[26] X. Zhou, R. Zafarani, Fake news: A survey of research and detection methods and opportunities (2018). arXiv:1812.00315.

[27] O. Oluwabunmi, G. Cosma, A. Liotta, Clan-based cultural algorithm for feature selection, in: International Conference on Data Mining Workshops (ICDMW), IEEE, 2019, pp. 465–472. doi:10.1109/ICDMW.2019.00073.

[28] A. L. Barabási, M. Pósfai, Network Science, Cambridge University Press, Cambridge, UK, 2016.

[29] V. Latora, V. Nicosia, G. Russo, Complex Networks: Principles, Methods and Applications, Cambridge University Press, Cambridge, UK, 2017.

[30] M. Newman, Networks: an introduction, Oxford University Press, 2010.

[31] G. Strang, Introduction to linear algebra, Vol. 3, Wellesley-Cambridge Press Wellesley, MA, 1993.

[32] I. E. Antoniou, E. T. Tsompa, Statistical analysis of weighted networks, Discrete dynamics in Nature and Society (2008). doi:10.1155/2008/375452.

[33] M. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Characterization and modeling of weighted networks, Physica A: Statistical Mechanics and its Applications 346 (1-2) (2005) 34–43. doi:10.1016/j.physa.2004.08.047.

[34] J. Saramäki, M. Kivelä, J. P. Onnela, K. Kaski, J. Kertész, Generalizations of the clustering coefficient to weighted complex networks, Phys Rev E 75 (2007) 027105. doi:10.1103/PhysRevE.75.027105.

[35] J. Travers, S. Milgram, The small world problem, Psychology Today 1 (1) (1967) 61–67.

[36] S. Milgram, An experimental study of the small world problem., Sociometry American Sociological Association 32 (4) (1969) 425–443. doi:10.2307/2786545.

[37] P. Erdős, A. Rényi, On Random Graphs I, Publicationes Mathematicae 6 (1959) 290–297.

[38] E. N. Gilbert, Random graphs, Annals of Mathematical Statistics 30 (4) (1959) 1141–1144. doi:10.1214/aoms/1177706098.

[39] D. Watts, S. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442. doi:10.1038/30918.

[40] A. L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512. doi:10.1126/science.286.5439.509.

[41] M. Sparrow, The application of network analysis to criminal intelligence: An assessment of the prospects, Social Networks 13 (3) (1991) 251–274. doi:10.1016/0378-8733(91)90008-H.

[42] P. Klerks, The network paradigm applied to criminal organisations: Theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands, Connections 24 (3) (2001) 53–65.

[43] L. Freeman, Centrality in social networks conceptual clarification, Social Networks 1 (3) (1978) 215–239. doi:10.1016/0378-8733(78)90021-7.

[44] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proc of the National Academy of Sciences 101 (11) (2004) 3747–3752. doi:10.1073/pnas.0400087101.

[45] J. E. Hirsch, An index to quantify an individual's scientific research output, Proc Natl Acad Sci USA 102 (46) (2005) 16569–16572. doi:10.1073/pnas.0507655102.

[46] L. Lü, T. Zhou, Q. M. Zhang, H. E. Stanley, The h-index of a network node and its relation to degree and coreness, Nature Communications 7 (2016) 10 168–10 175. doi:10.1038/ncomms10168.

[47] U. Brandes, On variants of shortest-path betweenness centrality and their generic computation, Social Networks 30 (2) (2008) 136–145. doi:10.1016/j.socnet.2007.11.001.

[48] P. Bonacich, Power and centrality: A family of measures, American Journal of Sociology 92 (5) (1987) 1170–1182. doi:10.1086/228631.

[49] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1) (1953) 39–43. doi:10.1007/BF02289026.

[50] M. Benzi, C. Klymko, On the limiting behavior of parameter-dependent network centrality measures, SIAM Journal on Matrix Analysis and Applications 36 (2) (2015) 686–706. doi:10.1137/130950550.

[51] P. De Meo, M. Levene, F. Messina, A. Provetti, A general centrality framework-based on node navigability, IEEE Transactions on Knowledge and Data Engineering 32 (11) (2020) 2088–2100. doi:10.1109/TKDE.2019.2947035.

[52] F. Morone, H. Makse, Influence maximization in complex networks through optimal percolation, Nature 524 (7563) (2015) 65–68. doi:10.1038/nature14604.

[53] C. W. Richard, Z. Ping, A study of graph spectra for comparing graphs and trees, Pattern Recognition 41(9) (2008) 2833–2841. doi:10.1016/j.patcog.2008.03.011.

[54] P. Wills, F. G. Meyer, Metrics for graph comparison: A practitioner's guide, PLoS ONE 15 (2) (2020) e0228728. doi:10.1371/journal.pone.0228728.

[55] C. Chen, H. Tong, B. Prakash, C. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, D. Chau, Node immunization on large graphs: Theory and algorithms, IEEE Transactions on Knowledge and Data Engineering 28 (1) (2016) 113–126. doi:10.1109/TKDE.2015.2465378.

[56] H. Hethcote, The mathematics of infectious diseases, SIAM review 42 (4) (2000) 599–653. doi:10.1137/S0036144500371907.

[57] J. Restrepo, E. Ott, B. Hunt, Characterizing the dynamical importance of network nodes and links, Physical review letters 97 (9) (2006) 094102. doi:10.1103/PhysRevLett.97.094102.

[58] H. Tong, B. Prakash, T. Eliassi-Rad, M. Faloutsos, C. Faloutsos, Gelling, and melting, large graphs by edge manipulation, in: Proc of the ACM international Conference on Information and Knowledge Management (CIKM 2012), ACM, Maui, 2012, pp. 245–254. doi:10.1145/2396761.2396795.

[59] P. Holme, B. Kim, C. Yoon, S. Han, Attack vulnerability of complex networks, Physical Review E 65 (2002) 5. doi:10.1103/PhysRevE.65.056109.

[60] S. Agreste, S. Catanese, P. De Meo, E. Ferrara, G. Fiumara, Network structure and resilience of mafia syndicates, Information Sciences 351 (2016) 30–47. doi:10.1016/j.ins.2016.02.027.

[61] B. Bollobás, Random graphs, Cambridge University Press, 2001.

[62] R. Albert, A. L. Barabási, Statistical mechanics of complex networks, Rev Mod Phys 74 (2002) 47–97. doi:10.1103/RevModPhys.74.47.

[63] Y. LeCun, Y. Bengio, G. Hinton, Deep Learning, Nature 521 (2015) 436–444. doi:10.1038/nature14539.

[64] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, Cambridge US, 2016.

[65] J. Stier, M. Granitzer, Structural analysis of sparse neural networks, Procedia Computer Science 159 (2019) 107–116. doi:10.1016/j.procs.2019.09.165.

[66] A. Bourely, J. P. Boueri, K. Choromonski, Sparse neural networks topologies, preprint (2017). arXiv:1706.05683.

[67] C. C. Hilgetag, A. Goulas, Is the brain really a small-world network?, Brain Structure and Function 221 (4) (2016) 2361–2366. doi:10.1007/s00429-015-1035-6.

[68] S. Liu, D. C. Mocanu, M. Arr, Y. Pei, M. Pechenizkiy, Sparse evolutionary Deep Learning with over one million artificial neurons on commodity hardware (2019). arXiv:1901.09181.

[69] G. Bellec, D. Kappel, W. Maass, R. Legenstein, Deep rewiring: Training very sparse deep networks, preprint (2018). arXiv:1711.05136.

[70] S. Srinivas, A. Subramanya, R. V. Babu, Training sparse neural networks, in: Proc of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, 2017, pp. 455–462.

[71] C. Louizos, M. Welling, D. P. Kingma, Learning sparse neural networks through l_0 regularization, preprint (2017). arXiv:1712.01312.

[72] J. Frankle, M. Carbin, The lottery ticket hypothesis: Finding sparse, trainable neural networks, preprint (2018). arXiv:1803.03635.

[73] J. O. Finckenauer, Problems of definition: What is organized crime?, Trends in Organized Crime 8 (3) (2005) 63–83. doi:10.1007/s12117-005-1038-4.

[74] F. Thrasher, The gang: A study of 1,313 gangs in Chicago, The University of Chicago Press, Chicago, US, 2013.

[75] P. Adler, Wheeling and and dealing: An ethnography of an upper-level drug dealing and smuggling community, Columbia University Press, New York, 1993.

[76] P. Reuter, Disorganized Crime - The Economics of the Visible Hand, MIT press Cambridge, MA (USA), 1983.

[77] D. Gambetta, The Sicilian Mafia The Business of Private Protection, Harvard University Press, Cambridge, 1996.

[78] P. Campana, Explaining criminal networks: Strategies and potential pitfalls, Methodological Innovations 9 (20597) (2016) 9911562274. doi:10.1177/2059799115622748.

[79] R. C. Van der Hulst, Introduction to social network analysis (sna) as an investigative tool, Trends in Organized Crime 12 (2) (2009) 101–121. doi:10.1007/s12117-008-9057-6.

[80] J. W. Johnsen, K. Franke, Identifying central individuals in organised criminal groups and underground marketplaces, in: Y. Shi, H. Fu, Y. Tian, V. V. Krzhizhanovskaya, M. H. Lees, J. Dongarra, P. M. A. Sloot (Eds.), Computational Science – ICCS 2018, Springer International Publishing, Cham, Switzerland, 2018, pp. 379–386. doi:10.1007/978-3-319-93713-7_31.

[81] F. Calderoni, S. Catanese, P. De Meo, A. Ficara, G. Fiumara, Robust link prediction in criminal networks: A case study of the Sicilian Mafia, Expert Systems with Applications 161 (2020) 113666. doi:10.1016/j.eswa.2020.113666.

[82] P. Campana, V. Federico, Cooperation in criminal organizations: Kinship and violence as credible commitments, Rationality and Society 25 (2013) 263–289. doi:10.1177/1043463113481202.

[83] P. Doreian, Doing social network research, Social Networks 43 (2015). doi:10.1016/j.socnet.2015.04.007.

[84] J. Xu, H. Chen, Criminal network analysis and visualization, Communications of the ACM 48 (6) (2005) 100–107. doi:10.1145/1064830.1064834.

[85] S. J. Strang, Network Analysis in Criminal Intelligence, Springer International Publishing, Cham, 2014. doi:10.1007/978-3-319-04147-6_1.

[86] P. Duijn, V. Kashirin, P. Sloot, The relative ineffectiveness of criminal network disruption, Scientific Reports 4 (4238) (2014) 1. doi:10.1038/srep04238.

[87] H. Chen, W. Chung, J. Xu, G. Wang, Y. Qin, M. Chau, Crime data mining: A general framework and some examples, IEEE Computer 37 (2004) 50–56. doi:10.1109/MC.2004.1297301.

[88] A. Bahulkar, B. K. Szymanski, N. O. Baycik, T. C. Sharkey, Community detection with edge augmentation in criminal networks, in: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2018, pp. 1168–1175. doi:10.1109/ASONAM.2018.8508326.

[89] G. Berlusconi, F. Calderoni, N. Parolini, M. Verani, C. Piccardi, Link prediction in criminal networks: A tool for criminal intelligence analysis, PLoS ONE 11 (4) (2016) 1–21. doi:10.1371/journal.pone.0154244.

[90] J. Sarnecki, Delinquent networks: Youth co-offending in Stockholm, Cambridge University Press, 2001. doi:10.1017/CBO9780511489310.

[91] C. Morselli, Career opportunities and network-based privileges in the cosa nostra, Crime, Law and Social Change 39 (4) (2003) 383–418. doi:10.1023/A:1024020609694.

[92] J. McGloin, Policy and intervention considerations of a network analysis of street gangs, Criminology and Public Policy 4 (3) (2005) 607–635. doi:10.1111/j.1745-9133.2005.00306.x.

[93] M. Natarajan, Understanding the structure of a large heroin distribution network: A quantitative analysis of qualitative data, Journal of Quantitative Criminology 22 (2) (2006) 171–192. doi:10.1007/s10940-006-9007-x.

[94] F. Calderoni, The structure of drug trafficking mafias: the 'Ndrangheta and cocaine, Crime, Law and Social Change 58 (3) (2012) 321–349. doi:10.1007/s10611-012-9387-9.

[95] F. Calderoni, E. Superchi, The nature of organized crime leadership: criminal leaders in meeting and wiretap networks, Crime Law Soc Change 72 (2019) 419–444. doi:10.1007/s10611-019-09829-6.

[96] G. Mastrobuoni, E. Patacchini, Organized crime networks: An application of network analysis techniques to the american mafia, Review of Network Economics 11 (2012) 3. doi:10.1515/1446-9022.1324.

[97] V. Krebs, Mapping networks of terrorist cells, Connections 24 (3) (2002) 43–52.

[98] F. Calderoni, Inside criminal networks, European Journal on Criminal Policy and Research 15 (1) (2010) 69–70. doi:10.1007/s10610-010-9118-7.

[99] P. Campana, F. Varese, Listening to the wire: criteria and techniques for the quantitative analysis of phone intercepts, Trends in Organized Crime 15 (1) (2012) 13–30. doi:10.1007/s12117-011-9131-3.

[100] S. Catanese, P. De Meo, E. Ferrara, G. Fiumara, Detecting criminal organizations in mobile phone networks, Expert Systems with Applications 41 (13) (2014) 5733–5750. doi:10.1016/j.eswa.2014.03.024.

[101] E. Ferrara, P. De Meo, S. Catanese, G. Fiumara, Visualizing criminal networks reconstructed from mobile phone records (2014). arXiv:1407.2837.

[102] D. M. Schwartz, T. D. Rouselle, Using social network analysis to target criminal networks, Trends in Organized Crime 12 (2) (2009) 188–207. doi:10.1007/s12117-008-9046-9.

[103] D. Cornish, The procedural analysis of offending and its relevance for situational prevention, Crime Prevention Studies 3 (1994) 151–196.

[104] G. Bruinsma, W. Bernasco, Criminal groups and transnational illegal markets, Crime, Law and Social Change 41 (1) (2004) 79–94. doi:10.1023/B:CRIS.0000015283.13923.aa.

[105] J. S. Coleman, Foundations of Social Theory, Belknap Press Series, Cambridge, MA, United States, 1990.

[106] K. Carley, J. S. Lee, D. Krackhardt, Destabilizing networks, Connections 24 (3) (2002) 79–92.

[107] N. Lin, K. S. Cook, R. S. Burt, Social Capital: Theory and Research, Transaction Publishers, 2001.

[108] C. Morselli, J. Roy, Brokerage qualifications in ringing operations, Criminology 46 (2008) 71–98. doi:10.1111/j.1745-9125.2008.00103.x.

[109] C. Morselli, Structuring Mr. Nice: Entrepreneurial opportunities and brokerage positioning in the cannabis trade, Crime, Law and Social Change 35 (3) (2001) 203–244. doi:10.1023/A:1011272411727.

[110] T. Spapens, Macro networks collectives, and business processes: An integrated approach to organized crime, European Journal of Crime, Criminal Law and Criminal Justice 18 (2) (2010) 185–215. doi:10.1163/157181710X12659830399653.

[111] D. Bright, C. Greenhill, T. Britz, A. Ritter, C. Morselli, Criminal network vulnerabilities and adaptations, Global Crime 18 (4) (2017) 424–441. doi:10.1080/17440572.2017.1377614.

[112] S. Sonnino, L. Franchetti, La Sicilia nel 1876, G. Barbèra, Firenze, Italy, 1877.

[113] E. Kleemans, C. De Poot, Criminal careers in organized crime and social opportunity structure, European Journal of Criminology 5 (1) (2008) 69–98. doi:10.1177/1477370807084225.

[114] E. Kleemans, H. Van de Bunt, The social embeddedness of organized crime, Transnational Organized Crime 5 (1999) 19–36.

[115] R. Sciarrone, L. Storti, The territorial expansion of mafia-type organized crime. the case of the italian mafia in germany, Crime, Law and Social Change 61 (1) (2014) 37–60. doi:10.1007/s10611-013-9473-7.

[116] M. Bouchard, On the resilience of illegal drug markets, Global Crime 8 (4) (2007) 325–344. doi:10.1080/17440570701739702.

[117] J. Ayling, Criminal organizations and resilience, International Journal of Law, Crime and Justice 37 (4) (2009) 182–196. doi:10.1016/j.ijlcj.2009.10.003.

[118] P. Williams, Transnational criminal networks. Networks and netwars: The future of terror, crime, and militancy (2001).

[119] B. McCarthy, J. Hagan, L. E. Cohen, Uncertainty cooperation, and crime: Understanding the decision to co-offend, Social Forces 77 (1) (1998) 155–184. doi:10.2307/3006013.

[120] R. Kinney, P. Crucitti, R. Albert, V. Latora, Modeling cascading failures in the north american power grid, The European Physical Journal B: Condensed Matter and Complex Systems 46 (1) (2005) 101–107.

[121] M. Chincoli, A. Liotta, Self-learning power control in wireless sensor networks, Sensors 2018 18 (2) (2018) 375. doi:10.3390/s18020375.

[122] A. Korn, A. Schubert, A. Telcs, Lobby index in networks, Physica A: Statistical Mechanics and its Applications 388 (11) (2009) 2221–2226. doi:10.1016/j.physa.2009.02.013.

[123] J. Bae, S. Kim, Identifying and ranking influential spreaders in complex networks by neighborhood coreness, Physica A: Statistical Mechanics and its Applications 395 (2014) 549–559. doi:10.1016/j.physa.2013.10.047.

[124] M. Kitsak, L. Gallos, S. Havlin, F. Liljeros, L. Muchnik, E. Stanley, H. Makse, Identification of influential spreaders in complex networks, Nature Physics 6 (11) (2010) 888. doi:10.1038/nphys1746.

[125] L. Cavallaro, A. Ficara, P. De Meo, G. Fiumara, S. Catanese, O. Bagdasar, W. Song, A. Liotta, Criminal network: The Sicilian Mafia. "Montagna Operation", Zenodo 0.0.1 (July 2020). doi:10.5281/zenodo.3938818.

[126] S. Villani, M. Mosca, M. Castiello, A virtuous combination of structural and skill analysis to defeat organized crime, Socio-Economic Planning Sciences 65 (C) (2019) 51–65. doi:10.1016/j.seps.2018.01.002.

[127] J. McGloin, H. Nguyen, The importance of studying co-offending networks for criminological theory and policy, Crime and networks (2014) 13–27.

[128] A. Rostami, H. Mondani, The complexity of crime network data: A case study of its consequences for crime control and the study of networks, PLoS ONE 10 (3) (2015) 1–20. doi:10.1371/journal.pone.0119309.

[129] D. Robinson, C. Scogings, The detection of criminal groups in real-world fused data: using the graph-mining algorithm "graphextract", Security Informatics 7 (1) (2018) 2. doi:10.1186/s13388-018-0031-9.

[130] L. Cavallaro, O. Bagdasar, P. De Meo, G. Fiumara, A. Liotta, Graph and network theory for the analysis of criminal networks, in: G. Fortino, A. Liotta, R. Gravina, A. Longheu (Eds.), Data Science and Internet of Things: Research and Applications at the Intersection of DS and IoT, Springer International Publishing, Cham, Switzerland, 2021, pp. 139–156. doi:10.1007/978-3-030-67197-6_8.

[131] T. Squartini, R. Mastrandrea, D. Garlaschelli, Unbiased sampling of network ensembles, New Journal of Physics 17 (2) (2015) 2, 023052. doi:10.1088/1367-2630/17/2/023052.

[132] T. P. Peixoto, Reconstructing networks with unknown and heterogeneous errors, Phys Rev X 8 (2018) 041011. doi:10.1103/PhysRevX.8.041011.

[133] M. Newman, Estimating network structure from unreliable measurements, Phys Rev E 98 (2018) 062321. doi:10.1103/PhysRevE.98.062321.

[134] S. Soundarajan, T. Eliassi-Rad, B. A. Gallagher, Guide to selecting a network similarity method, in: Proc of the 2014 SIAM International Conference on Data Mining (SDM), 2014, pp. 1037–1045. doi:10.1137/1.9781611973440.118.

[135] F. Emmert-Streib, M. Dehmer, Y. Shi, Fifty years of graph matching, network alignment and network comparison, Information Sciences 346-347 (2016) 180–197. doi:10.1016/j.ins.2016.01.074.

[136] C. Donnat, S. Holmes, Tracking network dynamics: A survey using graph distances, The Annals of Applied Statistics 12 (2) (2018) 971–1012. doi:10.1214/18-AOAS1176.

[137] M. Tantardini, F. Ieva, L. Tajoli, C. Piccardi, Comparing methods for comparing networks, Scientific Reports 9 (1) (2019) 17557. doi:10.1038/s41598-019-53708-y.

[138] L. Cavallaro, A. Ficara, F. Curreri, G. Fiumara, P. De Meo, O. Bagdasar, A. Liotta, Graph comparison and artificial models for simulating real criminal networks, in: R. Benito, C. Cherifi, H. Cherifi, E. Moro, L. M. Rocha, M. Sales-Pardo (Eds.), Complex Networks and Their Applications IX, Springer International Publishing, Cham, Switzerland, 2021, pp. 286–297. doi:10.1007/978-3-030-65351-4_23.

[139] F. Calderoni, C. Piccardi, Uncovering the structure of criminal organizations by community analysis: The infinito network, in: Tenth International Conference on Signal-Image Technology and Internet-Based Systems, 2014, pp. 301–308. doi:10.1109/SITIS.2014.20.

[140] F. Calderoni, Identifying mafia bosses from meeting attendance, in: A. Masys (Ed.), Networks and network analysis for defence and security, Springer International Publishing, Cham, Switzerland, 2014, pp. 27–48. doi:10.1007/978-3-319-04147-6_2.

[141] F. Calderoni, Predicting organized crime leaders, in: G. Bichler, A. Malm (Eds.), Disrupting Criminal Networks: Network Analysis in Crime Prevention, Lynne Rienner Publishers, 2015, pp. 89–110.

[142] F. Calderoni, D. Brunetto, C. Piccardi, Communities in criminal networks: A case study, Social Networks 48 (2017) 116–125. doi:10.1016/j.socnet.2016.08.003.

[143] R. Grassi, F. Calderoni, M. Bianchi, A. Torriero, Betweenness to assess leaders in criminal networks: New evidence using the dual projection approach, Social Networks 56 (2019) 23–32. doi:10.1016/j.socnet.2018.08.001.

[144] C. Piccardi, G. Berlusconi, F. Calderoni, N. Parolini, M. Verani, Oversize network (2016).

[145] A. Rostami, H. Mondani, Network complexity data (2015).

[146] L. M. Gerdes, K. Ringler, B. Autin, Assessing the abu sayyaf group's strategic and learning capacities, Studies in Conflict & Terrorism 37 (3) (2014) 267–293. doi:10.1080/1057610X.2014.872021.

[147] A. Ficara, G. Fiumara, P. De Meo, S. Catanese, Multilayer network analysis: The identification of key actors in a Sicilian Mafia operation, in: D. Perakovic, L. Knapcikova (Eds.), Future Access Enablers for Ubiquitous and Intelligent Infrastructures, Springer International Publishing, Cham, Switzerland, 2021, pp. 120–134. doi:10.1007/978-3-030-78459-1_9.

[148] S. Borgatti, M. Everett, L. Freeman, Ucinet for Windows: Software for social network analysis, Vol. 6, Analytic Technologies, Harvard, MA, 2002.

[149] L. Lü, D. Chen, X. Ren, Q. Zhang, Y. Zhang, T. Zhou, Vital nodes identification in complex networks, Physics Reports 650 (2016) 1–63. doi:10.1016/j.physrep.2016.06.007.

[150] P. De Meo, F. Messina, D. Rosaci, G. Sarnè, A. Vasilakos, Estimating graph robustness through the randic index, IEEE Transactions on Cybernetics 48 (11) (2018) 3232–3242. doi:10.1109/TCYB.2017.2763578.

[151] A. Logins, P. Karras, An experimental study on network immunization, in: Proc of the International Conference on Extending Database Technology, EDBT 2019, Lisbon, Portugal, 2019, pp. 726–729.

[152] C. Comin, L. da Fontoura Costa, Identifying the starting point of a spreading process in complex networks, Physical Review E 84 (5) (2011) 056105. doi:10.1103/PhysRevE.84.056105.

[153] D. Shah, T. Zaman, Rumors in a network: Who's the culprit?, IEEE Transactions on Information Theory 57 (2011) 5163–5181. doi:10.1109/TIT.2011.2158885.

[154] W. Dong, W. Zhang, C. Tan, Rooting out the rumor culprit from suspects, in: Proc of the IEEE International Symposium on Information Theory, IEEE, 2013, pp. 2671–2675. doi:10.1109/ISIT.2013.6620711.

[155] B. Prakash, J. Vreeken, C. Faloutsos, Efficiently spotting the starting points of an epidemic in a large graph, Knowledge and information systems 38 (1) (2014) 35–59. doi:10.1007/s10115-013-0671-5.

[156] D. Nguyen, P. Nguyen, M. Thai, Sources of misinformation in online social networks: Who to suspect?, in: Proc of the Military Communications Conference (MILCOM 2012), IEEE, 2012, pp. 1–6. doi:10.1109/MILCOM.2012.6415780.

[157] C. Budak, D. Agrawal, A. El Abbadi, Limiting the spread of misinformation in social networks, in: Proc of the International Conference on World Wide Web (WWW 2011), ACM, Hyderabad, India, 2011, pp. 665–674. doi:10.1145/1963405.1963499.

[158] S. P. Borgatti, K. M. Carley, D. Krackhardt, On the robustness of centrality measures under conditions of imperfect data, Social networks 28 (2) (2006) 124–136. doi:10.1016/j.socnet.2005.05.001.

[159] A. Ng, A. Zheng, M. Jordan, Link analysis, eigenvectors and stability, in: U. S. A. Seattle (Ed.), Proc of the International Joint Conference on Artificial Intelligence (IJCAI 2001), Lawrence Erlbaum Associates Ltd, 2001, pp. 903–910.

[160] A. W. Al-Dabbagh, Design of a wireless control system with unreliable nodes and communication links, IEEE Transaction on Cybernetics 49 (1) (2019) 315–327. doi:10.1109/TCYB.2017.2772869.

[161] J. Leskovec, J. Mcauley, Learning to discover social circles in ego networks (2012).

[162] B. Rozemberczki, R. Sarkar, Characteristic functions on graphs: Birds of a feather, from statistical descriptors to parametric models, in: Proc of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), ACM, 2020, p. 1325–1334. doi:10.1145/3340531.3411866.

[163] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: Densification and shrinking diameters, ACM Trans. Knowl. Discov. Data 1 (2) (2007) 2–es. doi:10.1145/1217299.1217301.

[164] J. Leskovec, K. J. Lang, A. Dasgupta, M. W. Mahoney, Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters, Internet Mathematics 6 (1) (2009) 29–123. doi:10.1080/15427951.2009.10129177.

[165] E. Cho, S. A. Myers, J. Leskovec, Friendship and mobility: User movement in location-based social networks, in: Proc of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, NY, USA, 2011, p. 1082–1090. doi:10.1145/2020408.2020579.

[166] J. McAuley, J. Leskovec, Image labeling on a network: Using social-network metadata for image classification, in: A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (Eds.), Computer Vision – ECCV 2012, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 828–841. doi:10.1007/978-3-642-33765-9_59.

[167] G. Stewart, J. Sun, Matrix perturbation theory (1990).

[168] C. Klymko, Centrality and communicability measures in complex networks: Analysis and algorithms, Ph.D. thesis, Emory University (2014).

[169] G. Mastrobuoni, The value of connections: Evidence from the italian-american mafia, The Economic Journal 125 (586) (2015) F256–F288. doi:10.1111/ecoj.12234.

[170] A. Gusrialdi, Z. Qu, S. Hirche, Distributed link removal using local estimation of network topology, IEEE Transactions on Network Science and Engineering 6 (3) (2019) 280–292. doi:10.1109/TNSE.2018.2813426.

[171] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, B. Kingsbury, Deep Neural Networks for Acoustic Modeling in Speech Recognition, IEEE Signal Processing Magazine 29 (2012) 82–97. doi:10.1109/MSP.2012.2205597.

[172] D. S. Berman, A. L. Buczak, J. S. Chavis, C. L. Corbett, A survey of Deep Learning Methods for Cyber Security, Information 4 (2019) 122. doi:10.3390/info10040122.

[173] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Communications of the ACM 60 (6) (2017) 84–90. doi:10.1145/3065386.

[174] Y. Dong, D. Li, Deep Learning and Its Applications to Signal and Information Processing [Exploratory DSP], IEEE Signal Processing Magazine 1 (2011) 145. doi:10.1109/MSP.2010.939038.

[175] C. Cao, F. Liu, H. Tan, D. Song, W. Shu, W. Li, Y. Zhou, X. Bo, Z. Xie, Deep Learning and Its Applications in Biomedicine, Genomics, Proteomics & Bioinformatics 16 (1) (2018) 17–32. doi:10.1016/j.gpb.2017.07.003.

[176] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, T. Blaschke, The rise of deep learning in drug discovery, Drug Discovery Today 23 (6) (2018) 1241–1250. doi:10.1016/j.drudis.2018.01.039.

[177] D. Ruano-Ordás, I. Yevseyeva, V. B. Fernandes, J. R. Méndez, M. T. M. Emmerich, Improving the drug discovery process by using multiple classifier systems, Expert Systems With Applications 121 (2019) 292–303. doi:10.1016/j.eswa.2018.12.032.

[178] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas USA, 2016, pp. 770–778.

[179] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Proc of the Annual Conference on Neural Information Processing Systems, Long Beach, USA, 2017, pp. 6000–6010.

[180] N. Kalchbrenner, E. Elsen, K. Simonyan, S. Noury, N. Casagrande, E. Lockhart, F. Stimberg, A. van den Oord, S. Dieleman, K. Kavukcuoglu, Efficient neural audio synthesis, in: J. Dy, A. Krause (Eds.), Proc of the 35th International Conference on Machine Learning, PMLR, 2018, pp. 2410–2419.

[181] J. Hestness, S. Narang, N. Ardalani, G. F. Diamos, H. Jun, H. Kianinejad, P. Mma, Y. Yang, Y. Zhou, Deep learning scaling is predictable, empirically, preprint (2017). arXiv:1712.00409.

[182] T. Gale, E. Elsen, S. Hooker, The state of sparsity in deep neural networks, preprint (2019). arXiv:1902.09574.

[183] K. Ullrich, E. Meeds, M. Welling, Soft weight-sharing for neural network compression, preprint (2017). arXiv:1702.04008.

[184] C. Haslinger, N. Schweifer, S. Stilgenbauer, H. Döhner, P. Lichter, N. Kraut, C. Stratowa, R. Abseher, Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status, Journal of Clinical Oncology 22 (19) (2004) 3937–49. doi:10.1200/JCO.2004.12.133.

[185] D. Cai, X. He, J. Han, T. S. Huang, Graph Regularized Non-negative Matrix Factorization for Data Representation, PAMI 33 (8) (2011) 1548–1560. doi:10.1109/TPAMI.2010.231.

[186] D. Cai, X. He, J. Han, Speed up kernel discriminant analysis, The VLDB Journal 20 (2011) 21–33. doi:10.1007/s00778-010-0189-3.

[187] F. Castaldo, Messina, arrestati il capo ed i sodali della "famiglia mafiosa di Mistretta", Grandangolo, il giornale di Agrigento 18 (January 2019).