UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

RICARDO LOPES DE ANDRADE

CENTRALITY METRICS: including nodes' attributes and generalizing geometric and homophily metrics in social network analysis

Recife
2021

RICARDO LOPES DE ANDRADE

CENTRALITY METRICS: including nodes' attributes and generalizing geometric and homophily metrics in social network analysis

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Engenharia de Produção.

Área de concentração: Pesquisa Operacional.

Orientador: Prof. Dr. Leandro Chaves Rêgo

Recife

2021

Catalogação na fonte:
Bibliotecária Sandra Maria Neri Santiago, CRB-4 / 1267

A554c Andrade, Ricardo Lopes de.

Centrality metrics: including nodes’ attributes and generalizing geometric and homophily metrics in social network analysis / Ricardo Lopes de Andrade. – 2021.

121 f.: il., figs., tabs., abrev. e siglas.

Orientador: Prof. Dr. Leandro Chaves Rêgo.

Tese (Doutorado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia da Produção. Recife, 2021.

Inclui referências.

1. Engenharia de produção. 2. Centralidade de média-p. 3. Centralidade disruptiva de energia. 4. Índice el generalizado. 5. Redes de comércio. 6. Redes de criminalidade. 7. Redes de coautoria. I. Rêgo, Leandro Chaves (Orientador). II. Título.

UFPE

658.5 CDD (22. ed.) BCTG/2022-320

RICARDO LOPES DE ANDRADE

CENTRALITY METRICS: including nodes' attributes and generalizing geometric and homophily metrics in social network analysis

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito parcial para a obtenção do título de Doutor em Engenharia de Produção. Área de concentração: Pesquisa Operacional.

Aprovada em: 21/12/2021.

BANCA EXAMINADORA

____
Prof. Dr. Leandro Chaves Rêgo (Orientador)
Universidade Federal de Pernambuco

____
Prof. Dr. Francisco de Sousa Ramos (Examinador Interno)
Universidade Federal de Pernambuco

____
Profa. Dra. Maísa Mendonça Silva (Examinador Interno)
Universidade Estadual de Campinas

____
Prof. Dr. José Antônio Fernandes de Macêdo (Examinador Externo)
Universidade de Federal do Ceará

____
Prof. Dr. Jesús Pascual Mena Chalco (Examinador Externo)
Universidade Federal do ABC

ABSTRACT

Many social interactions can be modeled by networks, in which social actors are represented by nodes and their relationships by edges. Researchers over the years have used social network analysis (SNA) to study the topological structure of the network and understand relational patterns. There are several centrality measures in the literature, with different criteria that define which nodes are more central. In addition, several studies seek to understand how network structures are formed. More recently, scholars have included in the SNA the actors' attributes in the search for a better understanding, given that these attributes can influence the way relationships occur and, consequently, affect the network structure. By exploring gaps in the literature, this PhD Dissertation aims to contribute to the advancement of this field proposing new metrics: (i) considering the geodesic paths among pairs of nodes, it is proposed a generalized measure, called p-means centrality, depending on the value given to the parameter p, several measures of centrality are obtained; (ii) two centrality measures, based on the law of gravity are proposed, in which the strength of the nodes' attributes is combined with the strength of the relationships between them, the main measure is called energy disruptive; (iii) to explore the network formation, two measures are proposed that generalize the EI Index, a measure of homophily. One can be applied in case of disjoint and non-disjoint groups and the other, more complete, also explores the case of fuzzy groups. Several networks were considered to test the use of these metrics: (i) the p-means centrality was applied in a co-authorship network and in a transport network; (ii) disruptive energy centrality was applied in two crime networks; (iii) the two measures that generalize the EI index were applied in a co-authorship network and in an international trade network. Among the best results obtained it can be highlighted: (i) the p-means centrality have shown that the most central nodes, defined by negative values of p, are closer to the nodes with the greatest spreading capacity, defined by the Susceptible-Infectious-Recovered (SIR) model; (ii) disruptive energy centrality, when used as a target method, was the most efficient strategy, providing greater network damage than other centrality measures analyzed; (iii) the EI index was able to explore the formation of networks in cases of non-disjoint groups and also fuzzy groups, proving the generalization of the measures. Therefore, the metrics proposed in this work enhance the SNA by unifying existing centrality metrics, incorporating nodes' attributes in SNA metrics, and extending the scope of homophily studies to more general types of groups.

Keywords: p-means centrality. energy disruptive centrality; generalized EI index; trade networks;

criminality networks; coauthorship networks.

RESUMO

Muitas interações sociais podem ser modeladas por redes, em que atores sociais são representados por nós e suas relações por arestas. Pesquisadores, ao longo dos anos, têm utilizado a análise de redes sociais (SNA, do termo em inglês) para estudar a estrutura topológica da rede e entender padrões relacionais. Existem várias medidas de centralidade na literatura, com critérios diferentes que definem quais nós são mais centrais. Além disso, diversos estudos buscam entender como se formam as estruturas das redes. Mais recentemente, estudiosos incluíram na SNA os atributos dos atores na busca por uma melhor compreensão, uma vez que esses atributos podem influenciar a forma como as relações ocorrem e, consequentemente, afetar a estrutura da rede. Esta tese de doutorado, ao explorar lacunas existentes na literatura, tem como objetivo contribuir para o avanço desse campo de estudo, propondo novas métricas: (i) considerando os caminhos geodésicos entre pares de nós, propõe-se uma medida generalizada, chamada de centralidade de média-p, dependendo do valor dado ao parâmetro p, várias medidas de centralidade são obtidas; (ii) propõe-se duas medidas de centralidade, baseadas na lei da gravidade, em que a força dos atributos dos nós é combinada com a força das relações entre eles, a medida principal é chamada de centralidade disruptiva de energia; (iii) para explorar a formação de redes, propõe-se duas medidas que generalizam o índice EI, uma medida de homofilia. Uma pode ser aplicado em caso de grupos disjuntos e não-disjuntos e a outra, mais completa, também explora o caso de grupos fuzzy. Consideramos diversas redes para testar o uso dessas métricas: (i) a centralidade de média-p foi aplicada em uma rede de coautoria e uma rede de transporte; (ii) a centralidade disruptiva de energia foi aplicada em duas redes de criminalidade; (iii) as duas medidas que generalizam o índice EI foram aplicadas em uma rede de coautoria e em uma rede de comércio internacional. Dentre os melhores resultados obtidos podemos destacar: (i) a centralidade de média-p mostrou que os nós mais centrais, definidos por valores negativos de p, estão mais próximos dos nós com maior capacidade de disseminação, definida pelo modelo SIR; a centralidade disruptiva de energia, quando utilizada como método de ataque, foi a estratégia mais eficiente, proporcionando maior dano à rede do que outras medidas de centralidade analisadas; (iii) o índice EI conseguiu explorar a formação das redes em casos de grupos não-disjuntos e também de grupos Fuzzy, comprovando a generalização das medidas. Portanto, as métricas propostas neste trabalho fortalecem a SNA unificando métricas de centralidade existentes, incorporando atributos dos nós em métricas de SNA e estendendo o escopo de estudos de homofilia para tipos mais gerais de grupos.

Palavras-chave: centralidade de média-p; centralidade disruptiva de energia; índice EI generalizado; redes de comércio; redes de criminalidade; redes de coautoria.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AA	American Airlines
AC	Andean Community
AG	Attribute Gravity
AST	Average Sentencing Time
BC	Betweenness Centrality
CAFTA-RD	Central America Free Trade Agreementand Dominican Republic
CAIS	Central American Integration System
CARICOM	Caribbean Community
CC	Closeness Centrality
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DC	Degree Centrality
EC	Eigenvector Centrality
ED	Energy Disruptive
EE	Economic Engineering
GM	Gravity Model
GPD	Gross Domestic Product
HC	Harmonic Centrality
LAIA	Latin American Integration Association
LE	Labor Engineering
LO	Logistics
NAFTA	North American Free Trade Agreement
NC	Network Capital

NE	Network Energy
OE	Operations Engineering
OECS	Organisation of Eastern Caribbean States
OR	Operational Research
OrE	Organizational Engineering
PE	Product Engineering
PR	PageRank
QE	Quality Engineering
RSL	Resource-Sharing Level
SCM	Southern Common Market
SE	Sustainability Engineering
SIR	Susceptible-Infected-Recovered
UA	United Airlines
UW	Unweighted
W	Weighted
WD	Weighted by days
WH	Weighted by hours
ZU	Z_unweighted
ZW	Z_weighted
SNA	Social Network Analysis

LIST OF SYMBOLS

G	Graph
$V(G)$	Set of nodes
v_i	Vertex i
$L(G)$	Set of edges
d	Geodesic distance
c	Path
C	Set of all paths
D_c	Degree centrality
e	Eccentricity
C_{Ecc}	Eccentricity centrality
C_c	Closeness centrality
C_H	Harmonic centrality
C_b	Betweenness centrality
n	Number of nodes in the graph
α	Probability of infection
β	Probability of recovery or death
M_p	Generalized mean
$x_i \dots x_n$	Set of positive real number
p	Non-zero real number
$N(v_i)$	Set of nodes

CONTENT

1 INTRODUCTION

A social network can be defined as a set of nodes (so-called actors), where each one of them has some sort of relationship (edges) to some or all other actors [1]. Both actors and relationships can be established in different ways, depending on the context. An actor can be a single person, a group, a company or even countries. The relationships can be a friendship relation between two people, a collaboration or the existence of a joint member between two teams, or a business relationship among companies [2].

Social network analysis (SNA) studies how the actors are embedded and act collectively within a structure, and how this structure can be systematically formulated [3]. This analysis is facilitated by the use of Graph Theory, through a set of mathematical algorithms [4]. Moreover, this area of study which is gaining further importance, with a growing number of publications on social influence, disease spread, information communication, citations, co-authorship, international trade, biology, transportation, safety and so on [5–10]. One of the major problems addressed by SNA scholars is the identification of the key players or influential spreaders in a social network and in the proposal of new methods and metrics of analysis [11,12], whose results have both theoretical and practical relevance [13–20]. Therefore, the study of centrality, an essential tool to identify the position of nodes in social networks, is the main object of study of social network analysis. In addition, because of the generality of the word “importance”, there are currently in the literature different centrality measures; each measure defines in different ways the centrality of a node. The most commonly used centrality measures are: degree centrality of a node which is the number of other nodes to which it is directly connected [21]; closeness centrality which is based on the distance between a node and every other node in the network [21]; betweenness centrality which measures the importance of a node in the connection of other nodes in the network [4, 21]; eigenvector centrality which is based on the idea that the prestige of a given node is related to the prestige of its neighbors [22]; and eccentricity which is the length of its longest geodesic path (the largest geodesic distance) to some other node in the network [4].

1.1 MOTIVATION AND JUSTIFICATION

It is observed that pioneers and current researchers have as their main focus the node’s centrality, i.e., they try to quantify the structural importance of the actors in the network [23]. In other words, SNA, for the most part, focuses heavily on the topological structure [14, 23, 24], that is, the set of relationships that individuals establish through their interactions with one another, analyzing patterns of interactions and implications [1,25]. Among the structural analysis metrics are cited eccentricity centrality, closeness centrality and the harmonic centrality. Known as geometric metrics, all of them are developed according to the geodesic paths between pairs of vertices [26] and are characterized by the notion that the closer, on average, a node is to the other nodes in the network, the more central it is. That is, all geometric measurements are influenced by distances to determine the centrality of nodes. The eccentricity centrality considers the least eccentric node is the most central. The closeness centrality considers the most central node is the one that is closest to the other nodes of the component, taking into account the arithmetic mean of the shortest paths. The Harmonic centrality also considers the most central node is the one that is closest to the other nodes, in the same component or not, taking into account the harmonic mean of the shortest paths. Although distance is an important factor in determining how influential the nodes are and can be applied in different ways, there is no general metric that can give a more broad view of the influence of these distances on the centralities of nodes. This is one gap in the literature that this work aims to fulfill.

It is common for measurements of centralities to disregard the characteristics or attributes of nodes. The studies that concentrate on the topological structure, analyze the social networks without considering the environment or the situations that the actors are inserted, i.e., the characteristics or attributes of the actors. Node attributes are not required for network definition and play no role in many network analytical studies. However, in some cases, node attributes can be very important in the analysis of networks, since it allows greater knowledge of the environment where the actors are inserted in. Nodes' attributes can affect the relational patterns of a social network, as it may change the nodes' sociability strategies, the way nodes influence and are influenced and also how individual behaviors are perceived [27]. Furthermore, according to Robins [23], to ignore the characteristics of the actors is to risk an incomplete, possibly contestable explanation, so individual attributes may be important for a fuller understanding of the social network.

Nevertheless, more recently, another front of researchers argues that key players or influential spreaders may be important for factors unrelated to centrality [15], such as: age, gender, ethnicity, skills and knowledge, psychological characteristics, social level, resources, among other individual factors. Therefore, some scholars aiming to improve the way of identifying critical actors in a network have proposed ways of combining the connections or ties between actors in a network, with human capital, information about the attributes (or resources) that an actor has (or can access). For example, in order to select the key actors in a criminal network, Schwartz and Rouselle [16] defined a measure that incorporates the strength of the actors (quantification of the attributes, called by “attribute weightings”, as well as the strength of the relationships that link the actors in the network, link weights. This measure is called Network Capital (NC). Lindelauf et al. [28] used the Shapley value (a measure applied in cooperative games) to find important nodes in a terrorist network, through the structure of the network and the resources of the nodes. The study of Bright et al. [15], using a drug trafficking network, combined the measures of degree, betweenness and attributes to give one numerical quantity, the Euclidean norm of the corresponding data point and showed that this integration increases efforts to identify the key actors. In a study of co-authorship networks, Liu et al. [24] developed a method that combined the nodes’ importance with the topological structure of the network. The measure changes the weight of the connections between the nodes by incorporating the nodes’ attribute weights. Applied to an international trade network, Andrade and Rêgo [14] developed a methodology that incorporates the individual attributes in the topological structure of the network, thus obtaining greater precision in the analysis of the nodes’ importance.

The attributes addressed above in the cited articles are numerical attributes that characterize the importance or influence of the actors, which are processed with prior information. For example, in Lindelauf et al. [28], the attributes of the terrorists were defined based on information such as affiliations and signs of radicalization. The study of Bright et al. [15] defined the trafficking member attributes based on the resources (tangible or intangible) they had. Liu et al. [24] defined the attributes based on the number of citations. In Andrade and Rêgo [14], the attributes of the nodes are the gross domestic product (GDP). Some studies, inspired by the law of gravity, proposed new measures of centrality [17, 18, 29, 30]. For this, the node centrality value is assigned as the node mass, given by one of the centrality measures. For example, in Ma et al.’s study [17], the k-shell value of each node is viewed as its mass, while Niu et al. [29] use as a node’s mass its degree centrality. The use of the nodes’ attribute weights and the measures inspired by the law of gravity were motivating to propose a new measure for identifying key nodes to be removed from the network to dismantle it.

Beyond that, these external factors to the network, non-numeric (qualitative factors), helps to increase the understanding of the patterns of interaction, what has proved to be useful in some cases. For example, some scholars are interested in knowing if the nodes of a network disproportionately establish links with others that are similar to them in some respect, i.e., they want to verify the occurrence of a higher incidence of relations between actors that have similar attributes. This phenomena is called in the literature attribute-based homophily [31]. Hence, homophily refers to the process of forming or eliminating links that occur with probability proportional to the similarity or dissimilarity of network actors. Therefore, it refers to one of many possible factors that determine the formation of the network structure [32]. For Yu et al. [33], it serves as an intuitive base to connect attribute and network topology. A measure widely used as a measure of homophily is the EI index, this measure essentially quantifies individuals' propensity to interact with similar actors [34, 35]. However, it has the limitation that it is only applicable to groups that form a partition of the set of actors in the network. Withal, disjoint groups rarely exist on a large scale in many empirical networks as actors can belong to many associative groups simultaneously, with various levels of affiliation [36]. This is another gap in the literature that this work tackles.

1.2 OBJECTIVES

The objectives can be identified by general objective and specific objectives.

1.2.1 General objective

This PhD Dissertation aims to contribute to the advancement of Social Network Analysis by proposing the use of new metrics. The proposed metrics fill some gaps in the literature by providing a general geometric centrality measure that extends some metrics commonly used in the literature, incorporating individual attributes into the identification of most influential nodes and extending a homophily analysis to overlapping and fuzzy groups.

1.2.2 Specific objectives

In order to achieve the general objective, some specific targets are defined:

a) to express the geometric measures using the generalized mean of the distances from all other nodes in the network to a given node;

b) develop a measure based on the law of gravity, using the actor's attribute weight as its mass and the distance of the shortest path between two actors as its distance;

c) formulate two new measures, one that quantifies the relational structure within and between groups that encompasses not only the analysis of disjoint groups, but also non-disjoint groups, and another that expands this measure to fuzzy groups;

d) apply the proposed metrics in real networks to demonstrate their usefulness and applicability.

1.3 CONTRIBUTIONS AND IMPACTS

Although social network analysis seems to refer to applications on websites or social networking application programs like Facebook, Twitter or Instagram, its application covers much larger areas that are not necessarily related to these, providing integrated services in the area of scientific research [37]. SNA is a powerful tool that assists decision-making by producing explanations for social phenomena in a wide range of disciplines, from psychology to economics. Thence, the results arising from SNA studies have potential benefit for the sectors of the social system. In this work, for example, the proposed measures were applied in different categories of networks: co-authorship, international trade, criminal, air and road routes.

Chapter 3 presents the proposed measure called p-means and was applied in two co-authorship networks, revealing the most influential researchers and consequently with greater capacity to disseminate information. A co-authorship network was also studied in Chapters 5 and 6, by the proposed generalized EI index measure, overlapping and fuzzy case, respectively. The results reveal whether researchers are influenced, in establishing connections, or not by the similarities presented. In this way, it contributes to the understanding of how co-authorship networks are formed. In extension, the results can help educational institutions and their researchers in decision-making regarding the creation and dissemination of knowledge.

The measures p-means and generalized EI Index were also applied in an international trade network. The results pointed which countries have greater trade flow capacity and whether their transactions are influenced by trade agreements or by the Human Development Index they present. Such results can contribute to the countries' strategic planning, regarding future trade agreements or even trade barriers.

Two criminal networks were studied in Chapter 4. In them, the proposed energy disruptive measure was applied, the measure was shown to outperform other measures, used as a comparison, in the disruption of networks. In this case, the measure can contribute to security decisions, outlining strategies to dismantle criminal networks. Building one of these criminal networks is also a contribution of this dissertation. It consists of a network of encounters of individual wearing anklet monitors and can be useful for other applications to prevent crimes.

In these applications, social and economic impacts are observed. However, due to the wide possibility of applications, the measures proposed in this work may impact other factors, such as the environmental and financial. The field of Industrial Engineering, for example, presents several types of problems that are solved by SNA. Varela et al. [38] focus on applying SNA to an industrial plant layout problem. Specifically, their work aimed to analyze how the use of SNA techniques is fundamental for the study of important relationships between entities in a manufacturing environment, such as jobs and resources, in the context of industrial plant layout analysis. Li and Lu [39] approaches the complex projects organization (CPO) as a dynamic and complex-related social network. This work "analyses social characters of the complex projects and CPO; identifies the CPO's social network factors; establishes the integrated organizational social network model (SNM) by using social network analysis (SNA) method". Sitko-Lutek et al.'s work [3] concluded that due to the flexibility of SNA applications, it can emerge as an important management tool in the areas of quality management. In fact, the work concluded that SNA is an important tool when applied to improve an operational process that requires two or more functions working together. Pigatto et al. [40] developed a new methodological framework by linking two methodologies to SNA and competitiveness analysis. The work showed that there is a positive relationship between the degree centrality and the level of competitiveness. The works cited emphasize the importance of the Social Networks methodology in Industrial Engineering. In the future, the measures proposed in this work may be applied to solve problems in different areas of knowledge, including the problems tackled in Industrial Engineering studies.

1.4 OUTLINE OF THE THESIS

Besides this introduction chapter, this PhD Dissertation has six additional chapters, briefly described as follows:

a) **Chapter 2** - Social Network Analysis is presented as an approach to investigate social structure. This chapter defines social networks and the main metrics for identifying influential nodes and understanding network formations.

b) **Chapter 3** - Following the trend of developing new measures that identify influential nodes in the network and based on traditional measures (topological structure), this chapter presents a measure called p-means centrality. This new measure has a parameter that manipulates the distances between the nodes producing results that vary between the degree centrality and eccentricity centrality. In addition to the traditional measures mentioned, the proposed measure manages to generate other centralities that successfully identify influential nodes in the network. This ability to determine influential nodes in the network was determined by the spreading capacity of the node analyzed by the SIR simulation model.

c) **Chapter 4** - In line with scholars who argue that the best way to identify influential nodes in the network is by combining external factors with factors inherent in the network, this chapter proposes some measures, based on the law of gravity. The measurements in this study also present a parameter that manipulates the distances between the nodes. The capacity to find influential nodes by this measure is given by the ability to select nodes that disrupt the network.

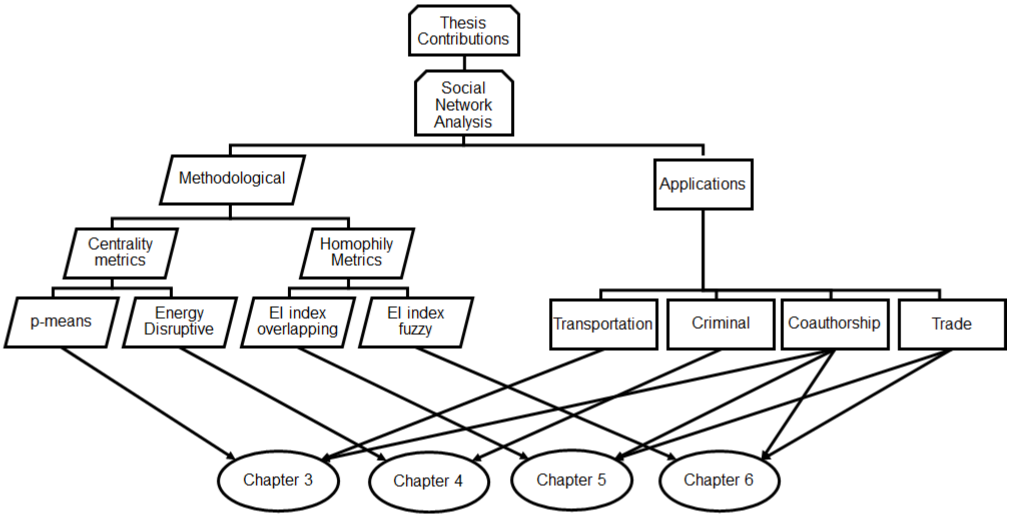
d) **Chapter 5** - The EI Index, a measure that analyzes relational patterns in the formation of networks, i.e., measures homophily, is the theme of this and the next chapter. In this chapter, it is attempted to generalize the EI Index to study the relational patterns in the presence of non-disjoint groups, given that, the current measure is limited to the analysis of disjoint groups.

e) **Chapter 6** - In this chapter an expansion of the previous proposal is made, through an adaptation seeking to include groups that present nodes with various levels of affiliations, fuzzy groups.

f) **Chapter 7** - The final chapter of this thesis summarizes the key achievements of the studies and their contributions to the social network analysis approach. It is also reported the limitations of the studies, the implications of their findings and recommendations for future research as well as others practical applications for the measures proposed here.

Figure 1 connects all the works developed, providing the visualization of the relationships among them.

Figure 1 – Relations among the works developed.



Source: The Author (2021)

Table 1 lists the metrics with the applications.

Table 1 – Relationship between metrics and applications

Contribution	Chapter	Transportation	Criminal	Coauthorship	Trade
p-means centrality	3	X	-	X	-
Energy disruptive	4	-	X	-	-
EI index overlapping	5	-	-	X	X
EI index fuzzy	6	-	-	X	X

Source: The Author (2021)

2 SOCIAL NETWORK ANALYSIS BACKGROUND

This chapter presents definitions and explanations about the fundamental topic of this work.

2.1 GENERAL DEFINITIONS

Mathematically, an undirected (or directed) and unweighted (or weighted) network has often been described as a simple graph $G = (V, L)$, where V is known as the set of nodes of the graph and L is the set of links or edges between the nodes of the graph.

In an unweighted network, all edges have the same strength, otherwise it is called weighted. If $n = ||V||$ is the cardinality of the set V, $v_i, v_j \in V$ and the graph is unweighted, A is the adjacent matrix $n \times n$, where $a(v_i, v_j)$, element of A, is equal to 1 if v_i is connected to v_j, or else $a(v_i, v_j)$ is set to be 0. In the case of a weighted network, we have that the weighted adjacent matrix of the graph is $W = w(v_i, v_j)_{nxn}$, where $w(v_i, v_j)$ is a non-negative weight of the edge between v_i and v_j. If there is no edge between v_i and v_j, then $w(v_i, v_j) = 0$.

Finally, an unweighted graph is called undirected if whenever $a(v_i, v_j) = 1$, then $a(v_j, v_i) = 1$. Otherwise, it is called directed. For a weighted graph to be undirected, it is necessary that for all nodes v_i and v_j, $w(v_i, v_j) = w(v_j, v_i)$. If this is not the case, then it is directed.

A *path* between two vertices v_i and v_j is a sequence of nodes $c = (v_0, v_1, v_2, \dots, v_k)$, where $v_0 = v_i$, $v_j = v_k$ and $a(v_{i-1}, v_i) = 1$, for $i = 1, 2, \dots, k$. Given a path $c = (v_0, v_1, v_2, \dots, v_k)$ between the vertices v_i and v_j, the length of this path is given by $d_c = k$. For undirected graphs, if there is a path between every pair of nodes in a graph, then the graph is called *connected*. A *component* of an undirected graph G is another undirected graph G_1 with the following properties: (i) $V(G_1) \subset V(G)$, (ii) $L(G_1) \subset L(G)$, (iii) G_1 is connected and (iv) there is no path in G connecting a node in $V(G_1)$ to a node in $V(G) - V(G_1)$. The component with the largest number of nodes in a graph is called the *giant component*.

A *geodesic path* is the shortest path between two vertices [41]. Thus, the geodesic path length, $d(v_i, v_j)$, also called *geodesic distance* or shortest distance, is the shortest distance in the network between these two vertices. Formally, given a path $c = (v_0, v_1, v_2, \cdots, v_k)$ between vertices v_i and v_j, the length of this path is given by $d_c = k$. Let $C(v_i, v_j)$ be the set of all paths between vertices v_i and v_j. Thus, if $C(v_i, v_j) \neq \emptyset$, then the geodesic distance is defined by:

$$d(v_i, v_j) = min\{d_c : c \in C(v_i, v_j)\}. \tag{1}$$

In the case of weighted networks, the length of a path $c = (v_0, v_1, v_2, \dots v_k)$ between vertices v_i and v_j, can be formally defined by Dijkstra's algorithm [1] and [42]: $d_c^w = \left(\frac{1}{w(v_0,v_1)} + \frac{1}{w(v_1,v_2)} + \cdots \frac{1}{w(v_{k-1},v_k)}\right)$. And the weighted geodesic distance is given by:

$$d^w(v_i, v_j) = min\{d_c^w : c \in C(v_i, v_j)\}. \tag{2}$$

The identification of the shortest paths and their length in directed networks is similar to the process in undirected networks. The single variation is that a path from one vertex to another can only follow the direction of the edges. In other words, the distance $d(v_i, v_j)$ from vertex v_i to vertex v_j in directed networks is the length of the shortest directed path from v_i to v_j. Getting distances from or to all other nodes, unlike in undirected networks, is relevant in the case of directed networks. For example, $d(v_i, v_j)$ is not necessarily equal to $d(v_j, v_i)$, in addition, it may occur that from v_j one can reach v_i but from v_i, v_j may not be reachable.

In the following subsections, it is recalled the definitions of some node centrality measures for weighted and unweighted networks.

2.2 DEGREE CENTRALITY

Proposed by Freeman[21], the degree centrality (DC) of node v_i, denoted by $C_d(v_i)$, is calculated by the number of nodes adjacent to vertex v_i. Formally, DC is defined by:

$$C_d(v_i) = \sum_{j=1}^{n} a(v_i, v_j). \tag{3}$$

The DC of node v_i, in a weighted network, is given by the sum of all the weights of the edges involving node v_i. For Newman [41] and Barrat et al. [43], the weighted DC (WDC) is given by:

$$C_d^w(v_i) = \sum_{j=1}^{n} w(v_i, v_j). \tag{4}$$

The DC is a simple and easy way to measure the local influence of a node [44, 45].

2.3 GEOMETRIC METRICS

The maximum distance from any vertex to vertex v_i in its same component is called *eccentricity*. Thus, if G_i is the component containing v_i, the eccentricity of v_i is given by:

$$e(v_i) = \max_{v_j \in V(G_i)} d(v_j, v_i). \tag{5}$$

The geometric metrics are those developed according to the geodesic paths between pairs of vertices [26]. Among the geometric metrics, besides the eccentricity, we highlight the *eccentricity centrality*, *closeness centrality* and the *harmonic centrality*. All geometric metrics are characterized by the notion that the closer, on average, a node is to the other nodes in the network, the more central it is.

The eccentricity centrality of vertex v_i [46], based on the eccentricity, is defined as:

$$C_{Ecc}(v_i) = \frac{1}{e(v_i)} \tag{6}$$

Consequently, the most important node in the network is the node with the smallest the longest geodesic path, in other words, the least eccentric vertex is the most central.

The *Closeness centrality* (CC) of a given node takes into consideration the geodesic distance from all other nodes in its same component to the given node. Freeman [21] asserted that the closeness centrality of vertex v_i, denoted by $C_c(v_i)$, is given by:

$$C_c(v_i) = \frac{1}{\sum_{v_j \in V(G_i)} d(v_i, v_j)}, \tag{7}$$

where G_i is the component containing v_i.

Larger values of $C_c(v_i)$ indicate smaller distances to the other nodes in the network, indicating that the node v_i takes an important position in the network. In weighted networks, the weighted closeness centrality (WCC) is given by:

$$C_c^w(v_i) = \frac{1}{\sum_j d^w(v_i, v_j)}. \tag{8}$$

The closeness centrality of a node measures its independence and efficiency in the communication with other nodes in the network [21].

The *Harmonic centrality* (HC) arose from a problem questioned by Marchiori and Latora [47] when analyzing networks which are not connected. Other authors also mention this same problem, such as Newman [48], Boldi and Vigna [26] and Wang [11]. To overcome the difficulty of forming a notion of "average shortest path", they proposed replacing the average (arithmetic) distance, used for the closeness centrality, by the harmonic mean of all distances. The formal definition of the normalized harmonic centrality, proposed by Rochat [49], is given by:

$$C_H(v_i) = \frac{1}{n-1} \sum_{i \neq j} \frac{1}{d(v_j, v_i)}. \tag{9}$$

For weighted networks, the definition is given as follows:

$$C_H^w(v_i) = \sum_{i \neq j} \frac{1}{d^w(v_i, v_j)}. \tag{10}$$

Another author Dekker [50], independently, also quotes this measure, called Valued Centrality. The author states that this measure is superior to the closeness centrality, which is "very sensitive to a single large distance or missing link".

2.4 BETWEENNESS CENTRALITY

The *Betweenness centrality* (BC) of vertex v_i, denoted by $C_b(v_i)$, is given by Freeman [21] and Wasserman and Faust [4]:

$$C_b(v_i) = \sum_{j,K} \frac{g(v_j, v_i, v_k)}{g(v_j, v_k)}. \tag{11}$$

where $g(v_j, v_k)$ is the number of shortest paths between vertices v_j and v_k and $g(v_j, v_i, v_k)$ is the number of shortest paths between vertices v_j and v_k going through vertex v_i.

In a weighted network, the BC is given by:

$$C_b^w(v_i) = \sum_{j,K} \frac{g^w(v_j, v_i, v_k)}{g^w(v_j, v_k)}. \tag{12}$$

where $g^w(v_j, v_k)$ is the number of weighted shortest paths between vertices v_j and v_k and $g^w(v_j, v_i, v_k)$ is the number of weighted shortest paths between vertices v_j and v_k going through vertex v_i, considering the weighted distance, $d^w(v_i, v_j)$.

The greater the BC of a node, the greater the capacity of the node to control the flow of information. According to Freeman [21]; Abbasi et al. [44], the BC is an indicator of the potential of a node to play the role of “mediator".

2.5 EIGENVECTOR CENTRALITY

A metric of importance of the node in the network based on its connections, the *Eigenvector centrality* (EC) is supported on the idea that a particular node will have high centrality if it is connected to vertices with central positions in the network [22]. In other words, the centrality of the vertex does not depend only on the number of adjacent vertices but also on the centrality of these vertices. Let λ be a constant, then the eigenvector centrality of node v_i, denoted by $C_e(v_i)$, is given by:

$$C_e(v_i) = \frac{1}{\lambda} \sum_{j=1}^{n} a(v_i, v_j) C_e(v_j). \tag{13}$$

Using the vector notation, let $X = (C_e(v_1), C_e(v_2) \ldots C_e(v_n))$ be the vector of eigenvector centralities. It can be rewritten Equation 13 as $\lambda X = AX$. By assuming that the eigenvector centrality takes only non-negative values (using the Perron-Frobenius theorem), it can be shown that λ is the largest eigenvalue of the adjacency matrix, where X is the corresponding eigenvector [51]. In the case of weighted networks, the adjacency matrix is replaced by the weighted adjacency [41]. And the eigenvector centrality is defined by:

$$C_e^w(v_i) = \frac{1}{\lambda} \sum_{j=1}^{n} w(v_i, v_j) C_e^w(v_j). \tag{14}$$

2.6 PAGERANK

PageRank (PR) is one metric used to rank web pages according to the interest and attention devoted to them [52]. PageRank takes into account the number and quality of links to a web page in order to determine how influential it is [24]. Let T_A be a web page and T_i be one of the web pages that has a link to T_A. Brin and Page [53] defined PageRank as follows:

$$PR(T_A) = (1 - \delta) + \delta \left(\frac{PR(T_1)}{C(T_1)} + \cdots + \frac{PR(T_n)}{C(T_n)} \right), \tag{15}$$

where $PR(*)$ is the PageRank of *, $C(T_i)$ is the number of links leaving page T_i and δ is a damping factor (if a person randomly clicks on pages and eventually stops clicking, then δ is the probability that, at any given moment, the person will continue to click), which can be chosen in the interval $(0, 1)$.

2.7 ATTRIBUTE WEIGHT

Liu et al. [24] and Andrade and Rêgo [14] developed methods to include nodes' attributes in the SNA. Thus, it is possible to classify networks as unweighted, weigthed by edges and weighted by edges and nodes. Since Liu et al.'s method transform the network into an asymmetric relation, the Andrade's method maintains the symmetric nature of network. For this reason, here is focused on this method.

The metric proposed by Andrade and Rêgo [14] as a way to take into consideration the importance of the node in the network context is given by:

$$Z(v_i, v_j) = w(v_i, v_j) \left(\frac{s(v_i) + s(v_j)}{2} \right), \tag{16}$$

where $Z(v_i, v_j)$ is equal to the original weight, $w(v_i, v_j)$, of the edge between the vertices, v_i and v_j, multiplied by the average of both nodes' attributes. The node's attributes are measurable characteristics associated with the type of relationship that connects them. With the incorporation of the node's attributes in the network, $Z(G)$ shall be the updated weighted adjacency matrix, where $Z(v_i, v_j)$ is an element of this matrix.

According to the authors, the inclusion of the nodes' weights contributes to a more efficient analysis of the network by combining factors inherent to the network with external factors.

2.8 HOMOPHILY and EI INDEX

Homophily is one of the most pervasive and robust trends in human interaction, describing how people tend to look for and interact with others that are more like them - often characterized as "birds of a feather", as named by Mcpherson et al. [32]. The studies on homophily in social networks, like those of [32, 54, 55], among others, seek to quantify the propensity of individuals to interact with similar actors. Furthermore, as a mechanism of social relations, it can explain the group composition in terms of social identities that range from ethnicity to age [56]. In fact, Mcpherson et al. [32] have argued that ethnicity, along with geography and kinship, are the main motivating factors behind homophily practices.

The EI index, proposed by Krackhardt and Stern [57], as a measure of homophily is essentially quantifying individuals' propensity to interact with similar actors [32, 54]. In this work, this measure quantifies the relational structure within and between groups [34, 35]. The EI index is an attractive measure of homophily because it does not depend on the density of a network [34]. The EI index is simply defined as the difference between ties between groups and intragroup ties, divided by the total number of ties for normalization.

$$EI \text{ index} = \frac{EL - IL}{EL + IL}, \tag{17}$$

where EL is the number of external links (links between nodes belonging to different groups); IL is the number of internal links (links between nodes belonging to the same group). The EI index ranges from -1 (all bonds are internal) to +1 (all bonds are external). The index can be calculated for the entire network, for each group or for each individual actor.

Everett and Borgatti [34] are among the researchers who treat the EI index as a measure of homophily and heterophily, where smaller values (internal connections) indicate greater homophily and higher values (external connections) indicate lower homophily or greater heterophily. The EI index was implemented in the popular social network analysis package [58] as a measure for homophily, which analyzes the tendency of people to connect with others similar to themselves, as well as insertion, i.e., how a node or group of nodes decides to connect to other nodes in a network [59].

The EI index also assists in the analysis of segregation in social networks [60, 61]. In fact, homophily is a natural explanation for segregation [62]. According to [63], a segregated network is one in which it is observed that the connections are concentrated on the vertices with similar characteristics, i.e., are calculated by comparing the number of links within groups to some measure of the baseline [61]. Segregation is also defined as the “unequal” distribution of two or more groups of people in different units or social positions [60].

Some authors like Andrade and Rêgo [14] and Danchev and Porter [64] also used the EI index in weighted networks. The EI index was calculated using the weight of the edges, this way EL is the sum of the edge weights that connect different cells of the partition and IL is the sum of the edge weights that connect actors of the same cell of the partition. The EI index for weighted networks, as well as unweighted ones, assumes values between -1 and +1. The weight of the edges, depending on how the relationships are established, usually represents the strength or intensity of the relationship. Therefore, when the EI index value approaches -1 it means that the internal relations are stronger or more intense. As the index approaches +1 it shows that external relations are stronger or more intense.

In this work, it be will also considered the nodes' weights (quantitative features) and insert it in the topological structure of the network. For this, it is used the method proposed by Andrade and Rêgo [14] (Equation 16). As it was shown, by this method, the weight of the edge is equal to the frequency or strength of the relationship between two nodes multiplied by the average nodes' weights. The intuition is that in cases where information about quantitative features of nodes are available, a link's weight must not only depend on the strength of the connection (original edge weight) but also on the average importance of the nodes being connected.

3 P-MEANS CENTRALITY

The proposed measure presented in this chapter was developed under this PhD Dissertation, and was previously published in the journal Communications in Nonlinear Science and Numerical Simulation (Andrade and Rêgo [13]).

3.1 CONTEXT

The closeness centrality and the harmonic centrality are geometric metrics, i.e., they are developed according to the geodesic path between pairs of nodes. Described in Rochat [49] as an alternative to the closeness centrality, the harmonic centrality seeks to include the unreachable nodes by calculating the sum of all the inverse of the shortest path distances from a specific node to all the other nodes. Thus, according to Marchiori and Latora [47], the harmonic mean performs better than the arithmetic mean. In fact, the normalized closeness centrality is the reciprocal of the arithmetic mean of the distances between a given node and every other node in the network, while the normalized harmonic centrality corresponds to the reciprocal of the harmonic mean of those distances [26]. Both the closeness centrality and the harmonic centrality capture the notion that the "closer" the node is, on average, to the other nodes in the network, the more central it is. Hence, the most central nodes are those that have the highest closeness centrality or harmonic centrality and can facilitate (or prevent) faster diffusion.

In mathematics, the arithmetic and harmonic means are special cases of the generalized mean, also known as Hölder's mean or power mean. By similarity, this work proposes to express the closeness centrality, harmonic centrality and eccentricity centrality using the generalized mean of the distances of all other nodes in the network to a given node. These measures differ only in the value of a parameter. However, other centrality measures can be defined by using some other values for the parameter, p. This generalized mean of the distances is called by *p-means centrality*.

It is possible to evaluate,with this proposed centrality measure, the most influential nodes in a network, controlling the parameter of the generalized mean. A recent paper, by Agneessens et al. [65], also proposed a generalized measure of centrality based on closeness. This measure generalizes the harmonic centrality equation, where a parameter assigns importance (weight) to the geodesic distances. The degree centrality and the harmonic centrality are two particular cases obtained when specific values for the parameter are chosen.

To evaluate the performance of this new measure, was used a classic model of transmission of diseases [66], the Susceptible-Infected-Recovered (SIR) model, comparing the nodes' classification, defined by the spread capacity, with the classification given by the proposed measure, for some values of p. It was also sought, in this work, to verify robustness and understandable the new measure of centrality through the axioms for centrality of [26]. Based on four real networks, the simulations show that the p-means centrality yields better results for negative values of the parameter p. The p-means centralities for some negative parameters are also the only ones that attend all the axioms for centrality.

The structure of this chapter is divided as follows: Section 3.2 discusses about the SIR model and the generalized mean. Section 3.3 proposes the p-means centrality and shows some of its relation with other centrality measures and study which axioms of centrality it satisfies. Section 3.4, tests the efficiency and feasibility of the proposed measure through simulations using data from four real networks. Section 3.5 presents a discussion about the choice of the parameter p. In Section 3.6, work is concluded presenting the last considerations of the study and proposals for future work.

3.2 BACKGROUND

In this section, the SIR model of disease transmission is presented, and the generalized average is discussed.

3.2.1 SIR model

The SIR model is a classic model of disease transmission, introduced by Kermack and McKendrick [67]. The acronym is derived from the three states or conditions that a network node may occupy: susceptible (capable of being infected), infected (capable of infecting) or recovered (no longer capable of infecting or being infected). The idea is to analyze the speed at which the disease spreads in the network and the proportion of infected nodes when the disease starts a given node. It is expected that the more central a node is, the higher are both the speed and the proportion of infected nodes. In the SIR model, at the beginning, a node to be tested is defined as an infected node, and at each step, each infected node will randomly infect susceptible neighbors with the scattering rate α. After infection, an infected node can be removed (killed or recovered with immunity) with probability β.

3.2.2 Generalized Mean

In this section, it is discussed about generalized mean following the definitions presented in Sheldon [68] and Borwein and Borwein [69]. Given a set of positive real numbers,

$x_1 \leq x_2 \leq \cdots \leq x_n$, its generalized mean, M_p, is defined as:

$$M_p = \left[\frac{1}{n}(x_1^p + x_2^p + \cdots + x_n^p)\right]^{\frac{1}{p}} \tag{18}$$

where p is a non-zero real number. The generalized mean is also known as Hölder mean or power mean and includes as special cases some classical means. For example, for $p = 1$, the generalized mean is equal to the arithmetic mean:

$$M_1 = \left[\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right]. \tag{19}$$

For $p = -1$, the generalized mean is equal to the harmonic mean

$$M_{-1} = \left[\frac{1}{n}(x_1^{-1} + x_2^{-1} + \cdots + x_n^{-1})\right]^{-1}. \tag{20}$$

Although the generalized mean is not defined for $p = 0$, its limit as p tends to zero is equal to the geometric mean:

$$M_0 = \lim_{p \to 0} \left[\frac{1}{n}(x_1^p + x_2^p + \cdots + x_n^p)\right]^{\frac{1}{p}} = (x_1 x_2 \cdots x_n)^{\frac{1}{n}}. \tag{21}$$

For $p = 2$, the generalized mean is equal to the quadratic mean:

$$M_2 = \left[\frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2)\right]^{\frac{1}{2}}. \tag{22}$$

Finally, for p tending to positive and negative infinity values, we have, respectively:

$$M_\infty = \lim_{p \to \infty} M_p(x_1, x_2, \cdots, x_n) = \max\{x_1, x_2, \cdots, x_n\} = x_n \tag{23}$$

and

$$M_{-\infty} = \lim_{p \to -\infty} M_p(x_1, x_2, \cdots, x_n) = \min\{x_1, x_2, \cdots, x_n\} = x_1. \tag{24}$$

In the next section, it is used the notion of generalized mean to define a more general centrality measure.

3.3 P-MEANS CENTRALITY

The equations of the closeness and harmonic centralities, as well as the eccentricity centrality, defined in the Section 2, are special cases of the following centrality measure, called p-means centrality:

$$p - means(v_i) = \left(\frac{\sum_{j:v_j \in N(v_i)} (d(v_j, v_i))^p}{n-1} \right)^{-\frac{1}{p}}, \tag{25}$$

where $p \neq 0$ and $N(v_i)$ is the set of nodes v_j such that there exists a path starting at v_j and finishing at v_i.

Although Equation 25 is not defined for $p = 0$, it is defined the 0-means centrality as the geometric mean as follows:

$$0 - means(v_i) = \left(\prod_{j:v_j \in N(v_i)} (d(v_j, v_i)) \right)^{-\frac{1}{n-1}}. \tag{26}$$

If $N(v_i) = V(G) - \{i\}$, then, as p tends to 0, it follows that:

$$\begin{aligned}
\lim_{p \to 0} p - means(v_i) = \\
\lim_{p \to 0} \left(\frac{\sum_{j:v_j \neq v_i} (d(v_j, v_i))^p}{n-1} \right)^{-\frac{1}{p}} = \\
\lim_{p \to 0} exp(ln(\left(\frac{\sum_{j:v_j \neq v_i} (d(v_j, v_i))^p}{n-1} \right)^{-\frac{1}{p}})) = \\
exp(\lim_{p \to 0} \frac{-ln(\left(\frac{\sum_{j:v_j \neq v_i} (d(v_j, v_i))^p}{n-1} \right))}{p}).
\end{aligned}$$

Thus, applying L'Hôpital's rule, it is gotten

$$\begin{aligned}
exp(\lim_{p \to 0} \frac{-1}{\sum_{j:v_j \neq v_i} (d(v_j, v_i))^p} \sum_{j:v_j \neq v_i} (d(v_j, v_i))^p ln(d(v_j, v_i))) = \\
exp(\frac{-1}{n-1} \sum_{j:v_j \neq v_i} ln(d(v_j, v_i))) = \\
exp(ln(\prod_{j:v_j \neq v_i} d(v_j, v_i))^{\frac{-1}{n-1}}) = \\
(\prod_{j:v_j \neq v_i} d(v_j, v_i))^{\frac{-1}{n-1}} = 0 - means(v_i).
\end{aligned}$$

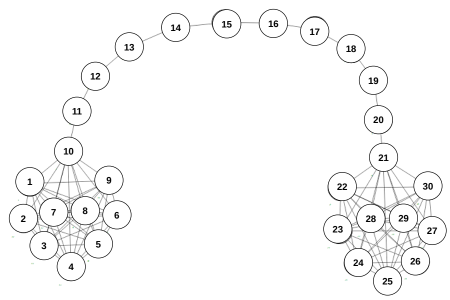
It is easy to see that for $p = 1$, we have that the p-means centrality is equal to the closeness centrality (Eq. 7), for $p = -1$ it is equal to the harmonic centrality (Eq. 9) and, as $p \to \infty$, it tends to the eccentricity centrality (Eq. 6).

Using Equations (23) and (24), it can be seen that the p-means centrality always lies between the reciprocal of the smallest and the reciprocal of the largest distance between every node in v_i's component and v_i. This is equivalent to the p-means centrality lying between the eccentricity centrality and 1. These extremes occur, respectively, for positive and negative infinity values of p.

Moreover, since distances between nodes are at least equal to one, it follows that as $p \to -\infty$, the term $d(v_j, v_i)^p$ goes to zero if $d(v_j, v_i) > 1$ and remains equal to 1, otherwise. Thus, as p becomes smaller, it follows that p-means(v_i) becomes closer to $(\frac{D_c(v_i)}{n-1})^{-\frac{1}{p}}$, which is an order preserving transformation of the degree centrality of node v_i. Therefore, the ranking generated by the degree centrality of a node is a particular case of the ranking generated by the p-means centrality of such node for small values of p. This result is illustrated by means of a Kendall correlation analysis in Subsection 3.4.4.

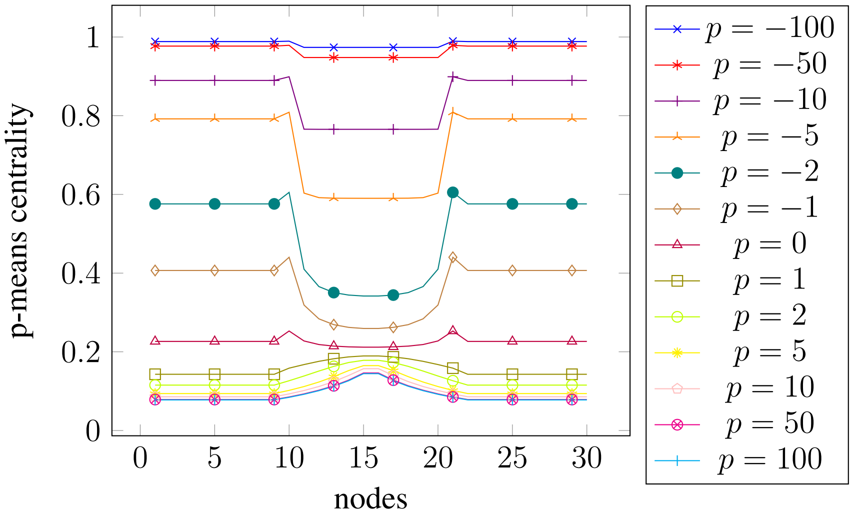
To illustrate the behavior of the proposed measure for different values of p, we present the network of Figure 2, which is based on the figure presented in [49]. The p-means centralities of these nodes are shown in Figure 3. For values of p ranging from -100 to 100, the p-means centrality ranges from approximately 1 to a value close to the eccentricity centrality, respectively. For negative p, the most connected nodes are the most influential. There is a mitigation of the impact of greater distances and an accentuation of the impact of small distances. On the other hand, for p positive, the greater distances strongly impact the definition of centrality and an expressive change of the influential nodes is observed.

Figure 2 – Example to illustrate the behavior of the p-means centrality.



Source: Andrade and Rêgo [13]

Figure 3 – p-means centralities of the nodes of the illustrative example, for different values of p.



Source: Andrade and Rêgo [13].

3.3.1 Axioms for centrality

Three axioms defined by Boldi and Vigna [26] are used to understand the behavior of the proposed measure. The axioms are to pave the way for a well-grounded assessment of centrality measures, analyzing robustness and understandability. The axioms are [26]:

(i) ***Definition 1 (Size axiom)*** *Consider the graph $S_{k,l}$ made by a k-clique and a directed l-cycle (Figure 4). A centrality measure satisfies the size axiom if for every k there is a P_k such that for all $l \geq L_k$ in $S_{k,l}$ the centrality of a node of the l-cycle is strictly larger than the centrality of a node of the k-clique, and if for every l there is a K_l such that for all $k \geq K_l$ in $S_{k,l}$ the centrality of a node of the k-clique is strictly larger than the centrality of a node of the l-cycle.*

(ii) ***Definition 2 (Density axiom)*** *Consider the graph $D_{k,l}$ made by a k-clique and a l-cycle $(l, k \geq 3)$ connected by a bidirectional bridge $x \leftrightarrow y$, where x is a node of the clique and y is a node of the cycle (Figure 5). A centrality measure satisfies the density axiom if for $k = l$ the centrality of x is strictly larger than the centrality of y.*

(iii) ***Definition 3 (Score Monotonicity axiom)*** *A centrality measure satisfies the score-monotonicity axiom if for every graph G and every pair of nodes x, y such that $(x, y) \notin L(G)$, when we add (x, y) to $L(G)$ the centrality of y increases. (In case of outgoing centrality metrics, this axiom should be modified to add (y, x) in $L(G)$ instead of (x, y).)*

It can be checked for which values of p, the p-means centrality satisfies the size axiom. Note that for any node x in the clique of $S_{k,l}$ its p-means centrality is given by:

$$\sqrt[p]{\frac{k+l-1}{k-1}},$$

for $p \neq 0$, and for $p = 0$, it is equal to 1.

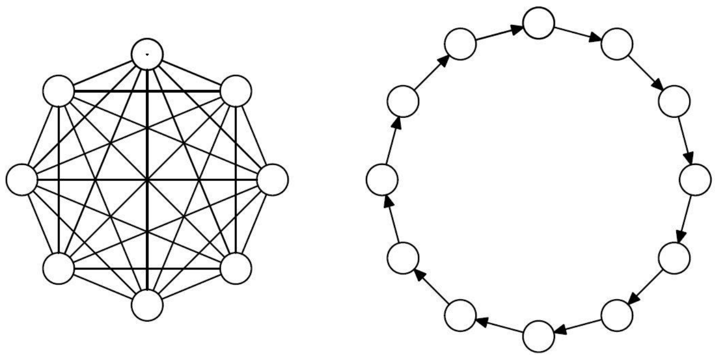
For any node y in the cycle of $S_{k,l}$ its p-means centrality is given by:

$$\sqrt[p]{\frac{k+l-1}{\sum_{i=1}^{l-1} i^p}},$$

for $p \neq 0$, and for $p = 0$, it is equal to $\frac{1}{\sqrt[k+l-1]{(l-1)!}} < 1$.

Therefore, as $k \to \infty$, the p-means centrality of x tends to 1, while the p-means centrality of y tends to 0 for $p < 0$ and to ∞ for $p > 0$. Moreover, as $l \to \infty$, the ratio of the p-means centralities of y and x tends to 0 for $p > 0$, tends to ∞ for $-1 \leq p < 0$ and to $\sqrt[p]{\frac{k-1}{\zeta(-p)}}$ for $p < -1$, where $\zeta(\cdot)$ is the Riemann function zeta. Thus, the size axiom is satisfied only for $-1 \leq p < 0$.

Figure 4 – Graph $S_{k,l}$.



Source:Boldi and Vigna [26].

It can be checked now for which values of p the p-means centrality satisfies the density axiom. Note that for the bridge node x in the clique of $D_{k,k}$ its p-means centrality is given by:

$$\sqrt[p]{\frac{2k-1}{k-1+\sum_{i=1}^{k} i^p}},$$

for $p \neq 0$, and for $p = 0$, it is equal to $\sqrt[1-2k]{k!}$.

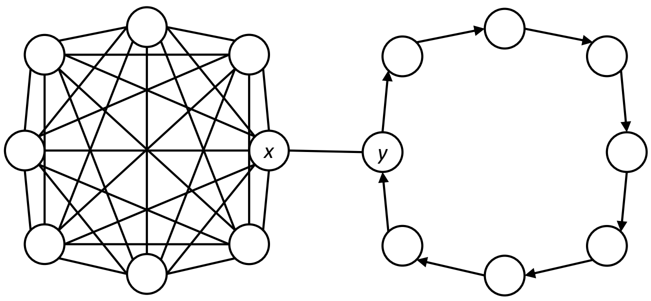
For the bridge node y in the cycle of $D_{k,k}$ its p-means centrality is given by:

$$\sqrt[p]{\frac{2k-1}{1+(k-1)2^p+\sum_{i=1}^{k-1} i^p}},$$

for $p \neq 0$, and for $p = 0$, it is equal to $\sqrt[1-2k]{2^{k-1}(k-1)!}$.

Therefore, as $k - 1 + k^p > 1 + (k-1)2^p$, for $k > 2$ and $p < 0$, we have that the p-means centrality of x is greater than the p-means centrality of y if $p < 0$. Since $\sqrt[1-2k]{k!} > \sqrt[1-2k]{2^{k-1}(k-1)!}$, we also have that the 0-means centrality satisfies the density axiom. On the other hand, for $k > 2$ and $0 < p < 1$, we have that $k - 1 + k^p < 1 + (k-1)2^p$. Thus, the p-means centrality of x is also greater than the p-means centrality of y if $0 < p < 1$. For $k > 2$ and $p \geq 1$, we have that $k - 1 + k^p \geq 1 + (k-1)2^p$ and the density axiom does not hold. Therefore, the density axiom holds for the p-means centrality if and only if $p < 1$.

Figure 5 – Graph $D_{k,k}$.



Source: Adapted from Boldi and Vigna [26].

Finally, it is checked now if the p-means centrality satisfies the Score Monotonicity axiom. Note that as we add a link from x to y we are either reducing the value of $d(x, y)$, if a path from x to y already was present in the graph or the node x became part of $N(y)$. In both cases, for $p < 0$, the summation $\sum_{j:v_j \in N(y)} (d(v_j, y))^p$ increases, what implies that the p-means centrality increases for $p < 0$. On the other hand, if $p \geq 0$ and a path from x to y was not present in G, then the p-means centrality may decrease ($p > 0$) or remain constant ($p = 0$). Thus, the score monotonicity axiom is only satisfied if $p < 0$.

Thus, the three axioms are only satisfied by the p-means centrality if $-1 \leq p < 0$.

3.4 EXPERIMENTAL ANALYSIS

In this section is tested the efficiency and feasibility of the proposed measure through simulations using data from four real networks.

3.4.1 Data

In order to analyze the performance of the proposed metric and how it changes for different values of the parameter p, we use data from four real networks. Next we give some details about these networks.

(i) USair97: It is a connected network with 322 airports, nodes, and the existence of a direct flight between two airports is denoted as a connection between these two nodes in the network, in the total of 2126 edges. The data can be downloaded from http://vlado.fmf.uni-lj.si/pub/networks/data/.

(ii) Netscience: This network is a representation of co-authorships between scientists working on network theory and experiments. Each node denotes a scientist, as compiled by M. Newman [70] and it is an undirected and unweighted network with 379 nodes in the giant component and 914 edges. The data can be obtained at http://www-personal.umich.edu/m̃ejn/netdata/.

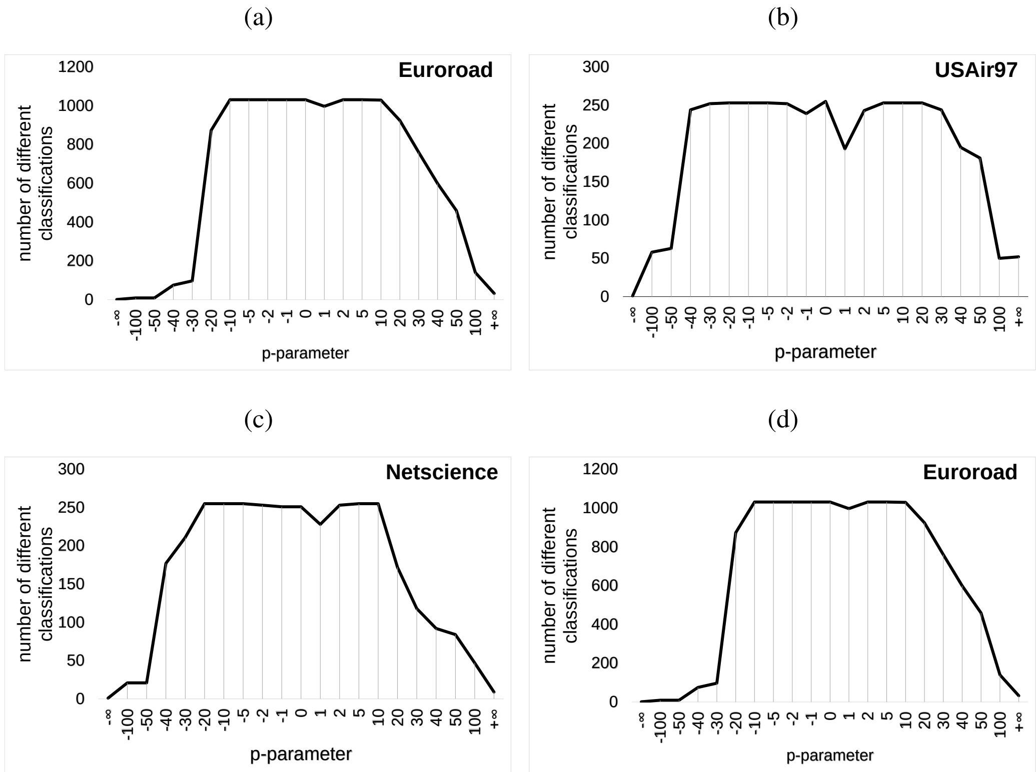
(iii) Euroroud: This is the international E-road network, a road network located mostly in Europe. The network is undirected, with 1039 nodes, in the giant component. The nodes represent cities and the edge denotes that they are connected by an E-road. There are 1035 edges. The data can be obtained at http://konect.uni-koblenz.de/publications.

(iv) Co-authorship PQ: The PQ network is co-authorship network among researchers in the area of Industrial Engineering of Brazil, has 92 nodes in the giant component and 131 edges. The network is undirected and the edges represent the publications in co-authorship [71].

3.4.2 p-Means Centrality Distinguishing Capacity

A good feature of any social network measure is its ability to distinguish the nodes of the network in the sense that there are as few ties as possible. Figure 6 seeks to analyze the distinguishing capacity of the p-means centrality for different p-values. For this analysis, after ranking the nodes we expose the last classification, because, in this way, the closer the last classification of the total number of nodes of the network, the greater the distinguishing capacity of the nodes. In the composition of the ranking, nodes that have the same centrality also have the same rating.

Figure 6 – The last classification of the node, for different values of p, in the four networks analyzed.



Source: Andrade and Rêgo [13].

The highest value for the last classification was obtained, in a common way in all networks, between $p = 10$ and $p = -10$. It can be noted that for $p = 1$, there is a loss in the ability to distinguish the nodes, represented in the graphs by a depression, which is less pronounced in the Euroroad network. The loss of the differentiation capacity of nodes is slower when p tends to positive infinity than when p tends to negative infinity. Since the former tends to the inverse of eccentricity and the latter tends to 1, that is, the shortest distance between two nodes.

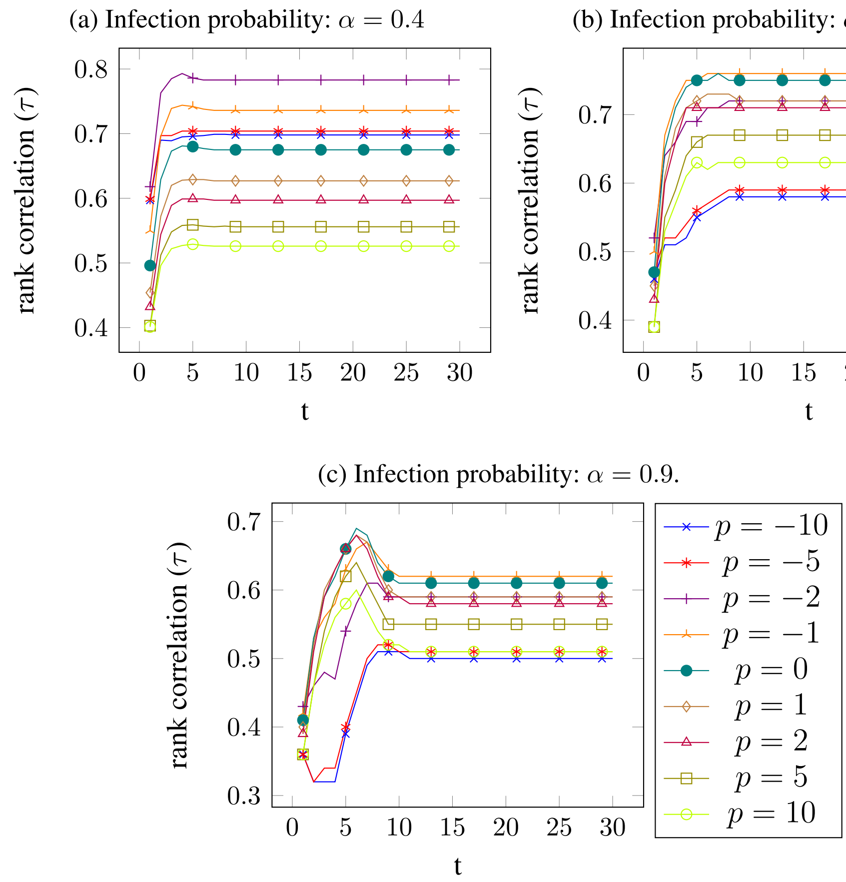
3.4.3 Effectiveness

In order to analyze the effectiveness of the p-means centrality, we follow a methodology different from the one commonly used in the literature, when compared to some model of transmission. For example Bian et al. [12] developed a method based on AHP and the effectiveness is checked when comparing with the W-TOPSIS method, by means of the spreading capacity of the top-10 nodes. Wang [11] proposed the efficiency centrality, whose performance was evaluated selecting the different top-10 nodes, generated by other centrality measures, to be the object of SIR model's simulation. Lu et al. [72] compared LeaderRank with PageRank by the spread capacity of top-20, top-50 and top-100 nodes generated by these metrics.

In our study, it is analyzed the spreading capacity of all nodes. Each node was initially infected at $t = 0$, the simulation stops when $t = 30$. At the end of the simulation, we checked the number of infected nodes $F(t = 30)$. Results are obtained by averaging 1000 implementations. Then, we ranked the nodes by this spreading ability. Finally, we compared this ranking with the raking generated by the p-means centrality for several values of p, through the Kendell correlation. In this way, the greater the correlation, the greater is the capacity of the p-means centrality to identify the nodes with greater spread capacity. By means of a previous analysis we selected three distinct probabilities of infection for simulating the SIR model, the criterion used was the ability to distinguish the nodes' centralities.

In Co-authorship PQ (Figure 7), we obtain results for $\alpha = 0.4$, $\alpha = 0.7$ and $\alpha = 0.9$. For $\alpha = 0.4$ (Figure 7a) we verified that the parameter $p = -2$ has the best approximation for the SIR model. Moreover, all other negative parameters values obtained greater correlation with the SIR method than their corresponding positive values. For $\alpha = 0.7$ (Figure 7b) we note that the parameter $p = -1$ has the best approximation for the SIR model. The parameter $p = -2$ also obtained a good result, but it was lower than the one obtained by $p = -1$. For $\alpha = 0.9$ (Figure 7c) we see that the parameter $p = -1$ remained with the highest correlation.

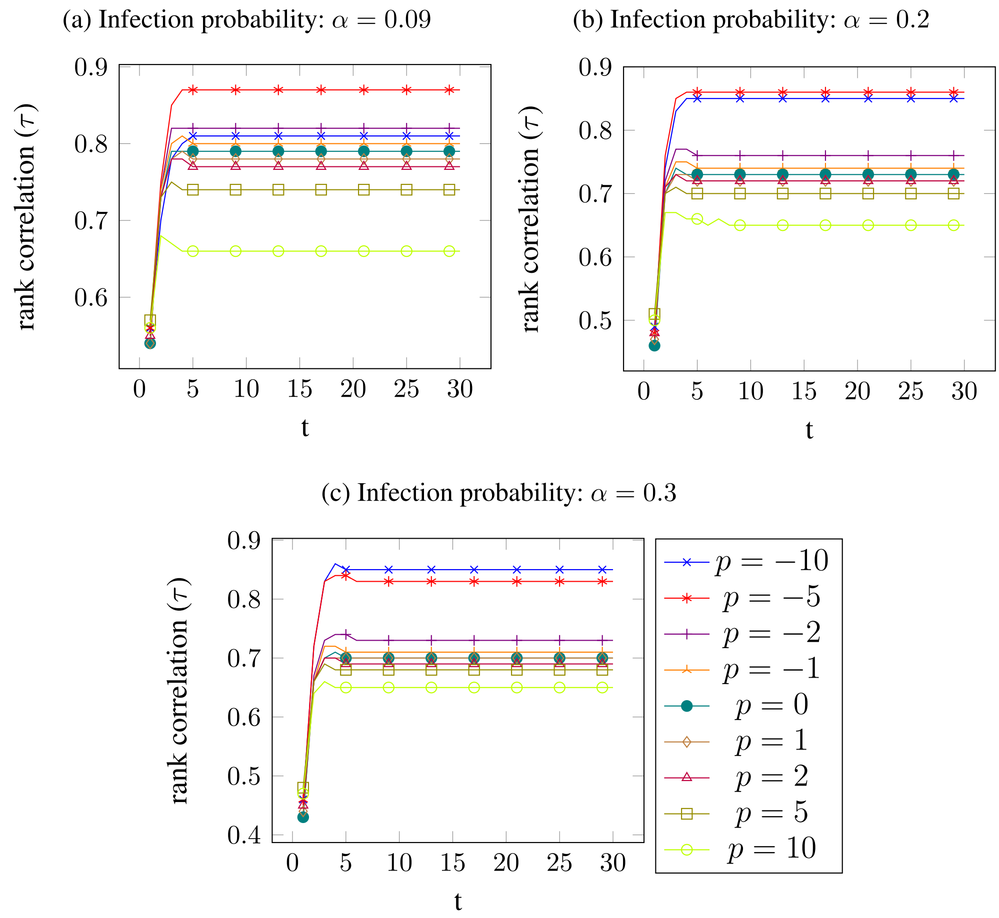
Figure 7 – In Co-authorship PQ, correlation between the classification generated by the SIR and the classification obtained by p-means centrality, for different values of p.



Source: Andrade and Rêgo [13].

In USair97 (Figure 8), we obtain results for $\alpha = 0.09$, $\alpha = 0.2$ and $\alpha = 0.3$. For $\alpha = 0.09$ (Figure 8a) we verified that the parameter $p = -5$ has the best approximation for the SIR model. For $\alpha = 0.2$ (Figure 8b), we note that the parameter $p = -5$ remained with the best approximation for the SIR model. For $\alpha = 0.3$ (Figure 8c), we see that the parameter $p = -10$ turned out to be the one with the highest correlation. In this network, no positive parameter valued obtained correlation that exceeded the one obtained by the corresponding negative value.

Figure 8 – In USair97, correlation between the classification generated by the SIR and the classification obtained by p-means centrality, for different values of p.

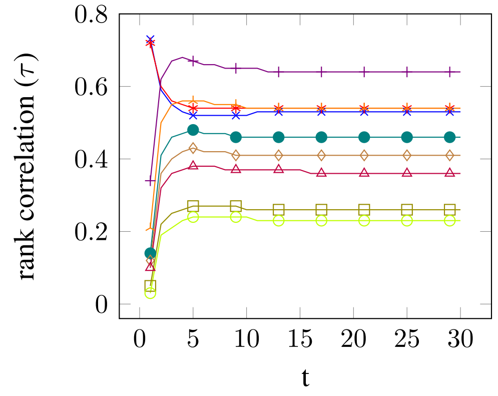


Source: Andrade and Rêgo [13].

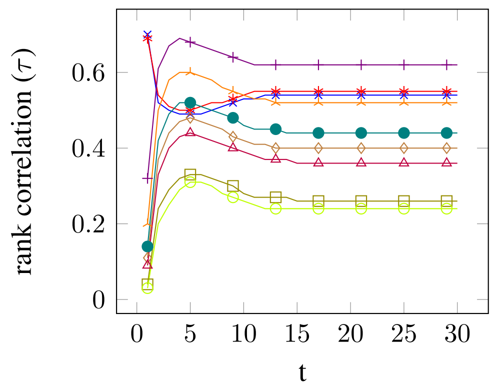
In Netscience (Figure 9), we obtain results for $\alpha = 0.3$, $\alpha = 0.4$ and $\alpha = 0.6$. For $\alpha = 0.3$ (Figure 9a), we verified that the parameter $p = -2$ has the best approximation for the SIR model. For $\alpha = 0.4$ (Figure 9b), we note that the parameter $p = -2$ remained with the best approximation for the SIR model. However, it had a slight drop, unlike the more negative parameter values that had a slight increase in their correlation with SIR. For $\alpha = 0.6$ (Figure 9c), we see that the parameters $p = -10$ and $p = -5$ obtained the best results. In this network, the positive parameters remained with lower correlations than their corresponding opposites.

Figure 9 – In Netscience, correlation between the classification generated by the SIR and the classification obtained by p-means centrality, for different values of p.

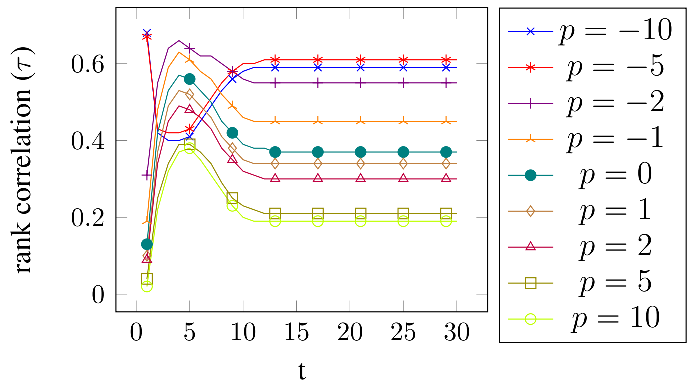
(a) Infection probability: $\alpha = 0.3$



(b) Infection probability: $\alpha = 0.4$



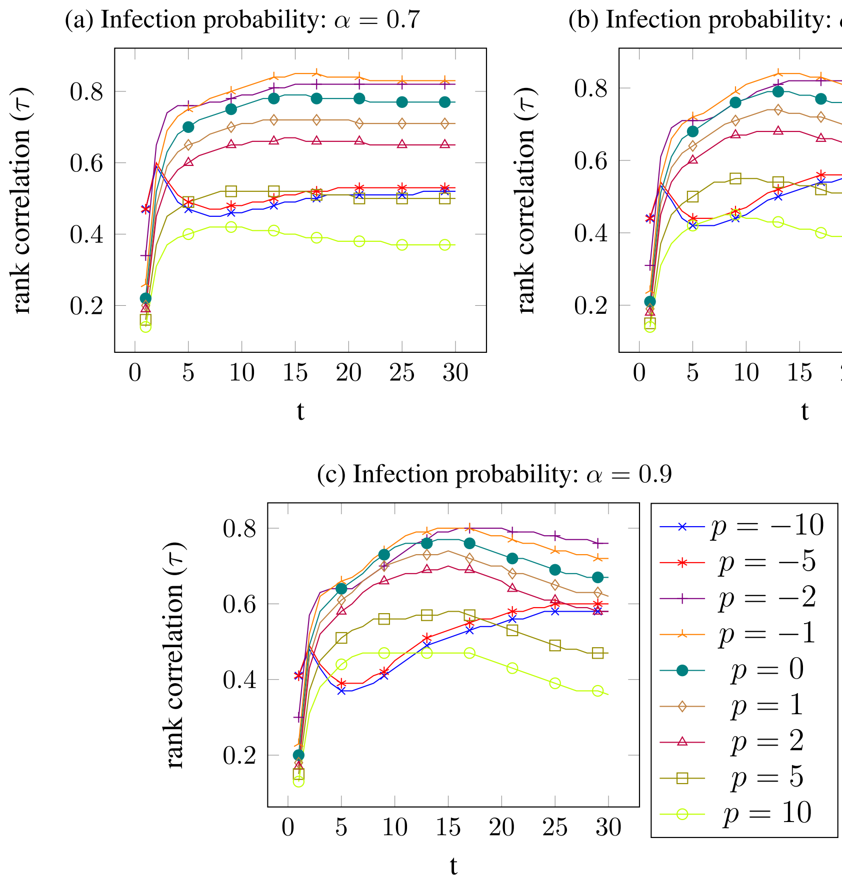
(c) Infection probability: $\alpha = 0.6$



Source: Andrade and Rêgo [13].

In Euroroad (Figure 10), we obtain results for $\alpha = 0.7$, $\alpha = 0.8$ and $\alpha = 0.9$. For all cases, $\alpha = 0.7$ (Figure 10a), $\alpha = 0.8$ (Figure 10b) and $\alpha = 0.9$ (Figure 10c), we note that the parameters $p = -1$ and $p = -2$ obtained similar results and have the best approximation for the SIR model.

Figure 10 – In Euroroad, correlation between the classification generated by the SIR and the classification obtained by p-means centrality, for different values of p.

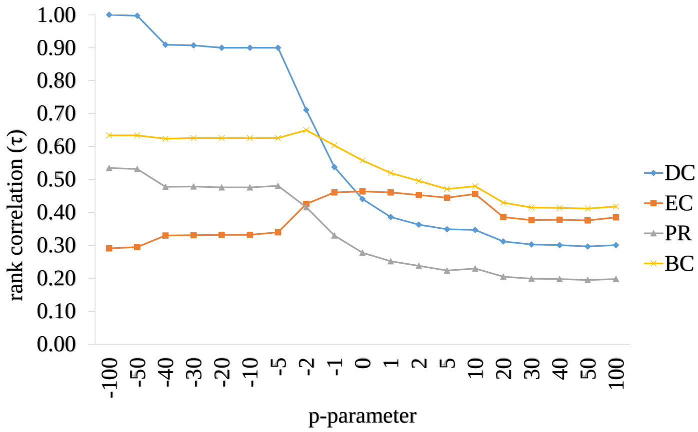


Source: Andrade and Rêgo [13].

3.4.4 p-means centrality and others centrality measures

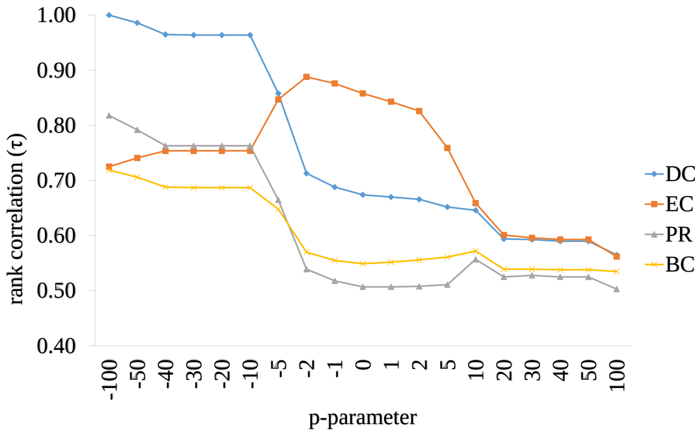
The p-means centrality was also compared with other centrality measures already well-known and used in the literature, namely: degree centrality (DC), betweenness centrality (BC), eigenvector centrality (EC) and PageRank (PR). Kendall correlations between p-means centrality and these measures for the four networks studied were generated.

Figure 11 – PQ, Kendall correlation between p-means centrality and others measures of centrality



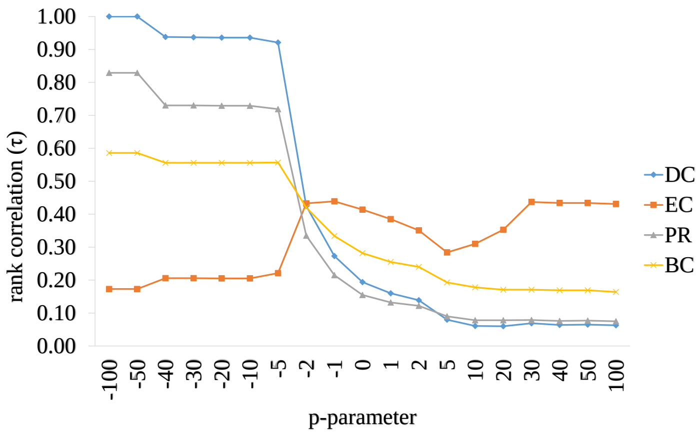
Source: Elaborated by the author (2020).

Figure 12 – USair97, Kendall correlation between p-means centrality and others measures of centrality



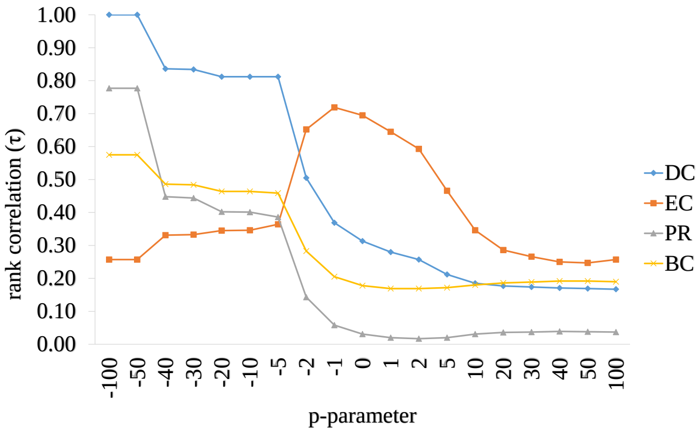
Source: Andrade and Rêgo [13].

Figure 13 – Netscience, Kendall correlation between p-means centrality and others measures of centrality



Source: Andrade and Rêgo [13].

Figure 14 – Euroroad, Kendall correlation between p-means centrality and others measures of centrality



Source: Andrade and Rêgo [13].

In the four networks analyzed, the same pattern of behavior is observed. The centralities DC, BC and PR have lower correlations with p-means centrality when p is positive and higher correlations when p is negative. The correlations between EC and p-means rise and fall, but are generally higher for positive values of p. The centralities DC, BC and PR show the highest correlation with the p-means centrality for p equal -100 and then remain stable for all p smaller than -100. As expected, as $p \to -\infty$, the Kendall correlation between DC and the p-means centrality goes to one.

3.5 THE CHOICE OF THE p-PARAMETER

It is proposed in this work a new centrality measure which depends on the choice of a parameter p. As shown in Figure 3, different values of p captures different centrality features. Indeed, it is showed that some commonly used centrality measures are special cases of the p-means centrality for some values of p, emphasizing that for different p values, the p-means centrality captures different properties of centrality. Moreover, Figure 6 shows that different values of p have distinct capacity of distinguishing the nodes' centralities. Therefore, as a general advice for practitioners, we suggest that the p-means centrality for a given network is analyzed for several values of p, as it was done in Figure 3 so that a more broad perspective about the nodes' centralities can be obtained.

However, our axiomatic study shows that the three axioms of centrality are only satisfied by the p-means centrality if $-1 \leq p < 0$. Thus, if these axioms are desired, one must restrict analysis to such range of values. From an empirical point of view, Figure 6 shows that the p-means centrality for values of p in the range $[-1, 0)$ also have a great capacity to distinguish the nodes' centralities. When we compare the results of the p-means centrality with that of the SIR model, we observe that higher correlations are obtained for non-positive values of p. Finally, in all networks analyzed, the p-means centrality obtained higher correlation with degree, betweeness and PageRank for negative values of p than for positive ones. These results provide some guidance on how to choose the values of p in a given network.

3.6 CONCLUSIONS

In conclusion, the experimental results on four real networks show that the negative p-parameters best detect the influential nodes, presenting a result closer to the SIR simulation model, that is, the p-means centrality for negative values of p obtains higher values for the nodes with greater spread capacity. The $p = 0$ also present better results than the positive p. The p-means centrality for negative p values also presented better results when it was compared the p-means centrality with the degree centrality, betweenness centrality and PageRank.

In this way, it is possible to conclude that the proposed model, which generalizes measures of centrality known in the literature and also generates other centrality, can successfully identify influential nodes in the networks. For values of p in the range $-1 \leq p < 0$, the p-means centrality satisfies the size, density and score monotonicity axioms for centrality, establishing itself as a robust and understandable identification of influential nodes.

4 ENERGY DISRUPTIVE CENTRALITY WITH AN APPLICATION TO CRIMINAL NETWORK

The proposed measure presented in this chapter was developed under this PhD Dissertation, and was previously published in the journal Communications in Nonlinear Science and Numerical Simulation (Andrade et al. [73]).

4.1 CONTEXT

Inspired by the law of gravity, some studies proposed new measures of centrality [17, 18, 29, 30]. In Ma et al. [17]'s study, the k-shell value of each node is viewed as its mass. K-shell is a decomposition method that divides the network into substructures (or levels) directly linked to centrality. Each node in the network is assigned a single integer index, k, according to the remaining degree, which is obtained by successively removing nodes with a degree less than the k_s value of the current layer [74, 75]. Niu et al. [29] use as a node's mass its degree centrality. Wang et al. [18] propose an improvement in the measure developed by Ma et al. [17]. Wang et al. [18] use the k-shell value of the focal node as its mass, while the DC, BC, CC or another centrality measure are considered as the mass of other neighboring nodes. The study concludes that combining K-shell with DC displays the best performance. Also based on Ma et al. [17]'s measure, Liu et al. [30] develop a variant algorithm called gravity model (GM), where the mass of the focal node and the destination node are both DC. All of these measures use as distance the shortest path length between the nodes, limited to a certain length.

In this work, it is proposed two gravity-law-based centrality measures and one global measure. In all of them, the actor's attribute weight is seen as its mass and the shortest path distance between two actors as its distance. For the first measure, Isaac Newton's idea of the gravity formula is used to measure the influence of the actors. This new measure is called by Attribute Gravity (AG) centrality. This measure quantifies the importance of the node by the sum of the force of attraction between it and every other node in the network. The second measure, called Energy Disruptive centrality, is based on how the removal of a node from the network changes the network energy, which is the proposed global measure. The network energy is the sum of the node's attributes and attribute gravity centralities. The core concept of this approach is the combination of two forces: attribute weight and the attraction force between nodes.

It is found in the literature the term "energy" in the definition of other network measures. The Laplacian energy of a network G is defined as the sum of absolute values of the eigenvalues of the Laplacian matrix of G [76]. Other measures have been defined for node ranking, for example, Average Energy Controllability Centrality [77] and Free-energy Rank [78]. The entropy, the internal energy and the Helmholtz free energy were also explored in several real-world networks in [79]. Although these metrics share the same name "energy" in their name, they are based on different concepts of energy and are not related to gravity forces as the one proposed in this work. Thus, since these metrics were designed with different purposes from this work, they will not be used.

Over the decade, some studies have addressed the analysis of social networks in combating criminal organizations. These studies address ways of disrupting criminal networks by eliminating the most important actors (given their position in the network or because they have fundamental resources for the functioning of the network) [15, 16, 80, 81] or characterize the fundamental factors for the working of criminal networks [82, 83]. The main studies address drug trafficking networks [15, 80], gangs [82, 83] and pedophilia networks [81].

In order to disrupt and dismantle a network both structurally and functionally, consequently avoiding possible criminal acts, this work compares the strategies of removing:

- the nodes with high centrality (degree, closeness, betweenness, pagerank and eigenvector);
- the most influential nodes according to the Euclidean distance method proposed by Bright et al. [15];
- the nodes whose removal reduces Network Capital (NC), measure proposed by Schwartz and Rouselle [16]; and
- the nodes whose removal leads to the highest loss of network energy, by a sequential removal process of one node at each step.

The comparison among these strategies occurred through robustness, attribute load and toughness, which measures the network loss in terms number of nodes, attribute weights and edge's weights, respectively. It is pointed out that the last two damage measures are being proposed in this work.

Two networks are considered in our analysis. The first one is formed by individuals sentenced to house arrest who wear an ankle monitor. The movement of criminals is an important factor applied in detecting crime [84, 85]. Ankle monitors movement data were used to establish connections between them and also to define the attribute weights. The second one is the operational network of hijackers of Al-Qaeda's 9/11 attack generated by Krebs [86] and the attributes are those determined by Lindelauf [28] based on information from the terrorists. The results demonstrate that energy disruptive centrality is the most efficient attack method - surpasses both measures that include node attributes and traditional measures - as it promotes the greatest damage to the network. The analysis, in the network of ankle monitors, also show that the removal of a single node using an efficient targeting measure causes more damage to the network than removing 20 convicts with greater recurrence in the crime of armed robbery. In the hijacker network, our measure can immediately target the hijackers who are links between hijacked aircraft groups.

It is argued that the proposed measures can be applied to analyze any type of network. For instance, in social media it can be interested in discovering who are the advertising influencers that engage and attract others who pay close attention to their product recommendation, posts and their views. The attributes in this case can be the number of likes in their posts, the number of comments and the number of buyers. By removing the network influencers, product marketing can run the risk of shrinking its potential reach and missing out on valuable customers.

The layout of this chapter is as follows. First, in Section 4.2, some traditional and recent measures used as a targeting method in the dismantling of networks are presented.. In Section 4.3, it is defined two proposed centrality measures and described how these will be used of them in this case work. Then the experimental results are presented in Section 4.4, where it is also proposed the attribute load and toughness metrics to measure network damage. Finally, discussions, conclusions and directions for future work are given in Section 4.5.

4.2 BACKGROUND ON DISRUPTIVE NETWORK

Methods based on nodes' centralities are widely discussed in the SNA literature [4, 21, 21, 21, 22, 53, 87–89]. Centered on the topological structure of the network, centrality measures seek to identify the most influential nodes in a given network structure, through a ranking. The better the position, the more influential the node is supposed to be. There are several centrality measures previously introduced by researchers, which include degree, betweenness, closeness, eigenvector and PageRank. All of these have been described in Chapter 2.

Networks, structures formed by nodes and edges, can be unstructured by removing nodes or removing edges, this process is referred to as percolation [90]. For Jahanpour and Chen [91], one of the great challenges today is to identify and remove important nodes in complex networks.

Some authors, Schwartz and Rouselle [16], Sparrow [92] and Carley et al. [93], who addressed network disruption and SNA, recommend targeting individuals with the highest centrality. In other words, targeting nodes based on their importance in the network's topological structure is a traditional form of network disruption. In addition, the best results are presented targeting nodes with the highest degree and betweenness centrality [93]. Identifying and targeting nodes that have unique functions or high cognitive load (in the literature it is usually called attributes or resources) are also ways to disrupt networks [93, 94]. Because of the difficulty of replacing these nodes, removing them will do its utmost to degrade the network's operational resources [95]. According to Carley et al. [93], an individual with a high cognitive load usually assumes more tasks, has more resources and communication/network ties. Characteristics inherent to leadership potential.

However, Borgatti [96], in a study that identifies key actors in a social network, shows that removing nodes with the highest traditional centrality measures (degree, closeness and betweenness) is not effective in disconnecting networks. Furthermore, the current measures of centrality are static, insomuch they consider only topological connectivity, disregarding the node's percolation state. Therefore, they are not suitable for this purpose [97]. Similarly, Carley et al. [93] also demonstrated that while there is often a correlation between an individual's position in the social network and their overall cognitive load, it is not perfect. Therefore, it is not clear whether there is a single node that can be removed to destabilize the network. Targeting the central node would not necessarily lead to a network break and that a targeted leader is not necessarily replaced by the most central actor. To really determine whether removing a node will destabilize a structure, it is needed to consider adaptation. Due to the individuals' ability to learn, the underlying social networks are dynamic.

As noted, an individual's cognitive load is influenced by social behavior, as communication/network ties are one of the factors that make up the cognitive load. However, individuals with a high cognitive load are not necessarily those with the highest centrality, but are equally critical to the network. Given these two aspects, cognitive load and centrality, an individual may have different positions: high cognitive load and high centrality, high cognitive load and low centrality, low cognitive load and high centrality and low cognitive load and low centrality. And the network must be seen not only as a topological structure, but also as a functional structure, a living organism, where exchanges of information, expertise, skills, knowledge occur, i.e., the sharing of the cognitive load. In this sense, individuals with high cognitive load and high centrality are the most important in the network, both functionally, they have what to share, and structurally, they have how to share. Therefore, targeting these individuals is the most effective way to destabilize the network. The other extreme is individuals who contribute little to the network, low cognitive load and low centrality, these are unlikely to be targeted. The problem occurs for individuals who present only one of the high aspects, as it is not clear which individual(s) to target. But the individual with the highest cognitive load or the highest centrality is usually targeted.

For the latter situation, some scholars argue that the best way to identify critical individuals is to combine SNA with the cognitive load of individuals [14–16, 24]. Schwartz [16] developed a measure, named "network capital" (NC), that incorporates the actors' strength or cognitive load, called by the authors "attribute weighting", as well as the strength of the relationships that link the actors in the network, link weights. Through this measure, the targeted individuals are those who contribute the most to the network capital, i.e., they are the ones who have a larger number of resources and who share them with other actors in the network. The NC is given by:

$$NC = \frac{Node\_Scores + Connection\_Scores}{n + [n(n-1)RSL]}, \tag{27}$$

where $Node\_Scores$ denotes the resources of the actors (attribute weighting); $Connection\_Scores$ denotes the way resources are shared, the value added by each actor to NC; n denotes the total number of actors in the network; and RSL can vary between 0 and 1, is a way to restrict resource sharing.

Bright et al. [15] in their study combined two measures of centrality (degree and betweenness) with the attribute weight to give a numerical quantity through the Euclidean distance. In a three-dimensional coordinate system, nodes are plotted using three values: degree centrality (x-axis), betweenness centrality (y-axis) and attribute weight (z-axis). Each axis varies between 0 and 1 inclusive, so the values, centralities and attributes, must be normalized, dividing by the maximum value of each measure. The numerical quantity of the combination of the three measures occurred in two ways: (1) Euclidean distance A, corresponds to the straight line distance from the origin (0, 0, 0) to the data point in three-dimensional space; (2) Euclidean distance B, corresponds to the distance of each data point from the corner of the cube (1, 0, 0). According to the authors, in criminal networks, the less visible actors (less likely to be targeted) can be identified through the trade-off between the degree centrality and the betweenness centrality. In addition, a strategic position, called brokerage positions [98], is obtained when the actors have low degree centrality and high betweenness centrality, through the Euclidean distance B.

4.3 NEW CENTRALITY MEASURES

Seeking a more efficient form of identifying influential nodes, involving topological and functional information and inspired by the law of gravity, this work proposes two centrality measures, called attribute gravity centrality (AG) and energy disruptive centrality (ED), and a global measure, called network energy (NE).

4.3.1 Attribute Gravity Centrality

Nodes with a high cognitive load (or attribute weight) are likely to be more influential. In addition, a node closer to the other nodes has greater influence. According to the above aspects, two nodes, whose masses are the attributes weights, are subject to mutual cooperation (sharing resources) proportional to their masses and inversely proportional to the square of the distance separating them. Thus, the AG of a node measures the sum of its mutual cooperation with all other nodes in the network and is given by:

$$AG_t(i) = \sum_{j \neq i} \frac{k_i k_j}{(d^w(v_i, v_j))^{1/t}}. \tag{28}$$

where k_i and k_j are the attributes weight of nodes i and j, respectively, $d^w(v_i, v_j)$ is the shortest distance between nodes v_i and v_j. (In unweighted networks replace $d^w(v_i, v_j)$ by $d(v_i, v_j)$.)

When there is no path between v_i and v_j , it is assumed $d^w(v_i, v_j) = +\infty$, consequently $AG_t(i, j) = 0$. The parameter t acts by changing the attractiveness of nodes and varies from zero to infinity. For t between 0 and 1, the distance between two nodes is increased, making the attraction force between the nodes smaller. The lower the value of t, the lower the force of attraction between distant nodes, leading to a devaluation of peripheral nodes. For t greater than 1, there is a shortening of the distances, the peripheral nodes become closer to the central nodes, thus increasing their interaction strength. The higher the value of t, the more homogeneous the distances are, to the point that the distances become insignificant. Therefore, the force of attraction between two nodes is increasingly independent of the distance and dependent on the attribute weights.

The AG, being developed according to the geodesic paths between pairs of vertices, entered into a family of measures known as geometric measures. The eccentricity centrality, closeness centrality, harmonic centrality and p-means centrality are geometric measures [13]. They are characterized by the notion that the closer, on average, a node is to the other nodes in the network, the more central it is. The propose measure is more similar to the harmonic centrality proposed by Rochat [49], defined as the harmonic mean of all distances. The harmonic centrality is superior to the proximity centrality because it is more sensitive to peripheral nodes [50]. The closeness, harmonic and the eccentricity centrality are special cases of the p-means centrality [13].

Although AG_t can be used to rank nodes in a network, it is not suitable to measure the impact of the removal of a node from the network since when a node is removed from a network not only its own AG_t is affected, but also of the other nodes which are in its same component. For that, in the next subsection, it is proposed a global measure which includes not only the force of attraction between nodes, but also the nodes' attributes.

4.3.2 Network Energy

The NE is the sum of the nodes' attribute weights and the attribute gravity centralities. The NE_t of a graph G is given as follows:

$$NE_t(G) = \sum_{i=1}^{n}(k_i + AG_t(i)). \tag{29}$$

The NE_t takes into account the total amount of resources at disposal of each node and the force of attraction between nodes.

It is possible now investigate what is the effect in the NE_t when t tends to extreme values 0 and infinity. First, when t tends to zero, $AG_t(i)$ tends to 0. Thus

$$NE_0(G) = \lim_{t\to 0} NE_t(G) = \sum_{i=1}^{n} k_i. \tag{30}$$

Second, when t tends to infinity, then:

$$NE_\infty(G) = \lim_{t\to\infty} NE_t(G) = \sum_{i=1}^{n}(k_i + \sum_{j:j\neq i, j\in G(i)} k_i k_j), \tag{31}$$

where $G(i)$ is the network component that contains node i.

Thus, for small values of t, each node contributes to $NE_t(G)$ according to its own attribute value. On the other hand, when t is very high the distance between nodes in a component becomes irrelevant and each node i besides contributing to $NE_t(G)$ with its own attribute value (k_i), also adds values to the network energy due to the synergy of exchanging attributes with other nodes in its same component ($k_i k_j$). For middle range values of t, this synergy is higher for closer nodes than for very distant nodes in the component. Thus, to dismantle the network, middle to high values of t should have a greater impact.

In the next section, it is proposed another centrality measure which ranks nodes according to the impact of their removal have on the network energy.

4.3.3 Energy Disruptive Centrality

Through $NE_t(G)$ it is possible to assess the importance of an isolated node in the graph G. When a node k is removed, the graph G is changed to a subgraph G'_k. The ED centrality of node k is defined as the drop in the NE caused by the removal of the node k from initial graph G. Formally,

$$ED_t(G, k) = NE_t(G) - NE_t(G'_k). \tag{32}$$

Thus, $ED_t(G, k)$ ranks nodes taking into account their attributes and the impact that their removal has on the synergy of attributes' exchange between nodes.

In the next section, it is studied the impact caused on a real network by sequentially removing nodes from a network one by one. If G^l is the subgraph of G after the removal of l nodes, it is proposed that the $l + 1$-th node to be removed should be the node k which maximizes $ED_t(G^l, k)$.

4.4 EXPERIMENTAL RESULTS

In this section, it is presented how the network of criminals encounters and the network of hijackers in the September 11 attacks were developed, and it is shown some network's global characteristics. It is also clarified how the attribute of the nodes was defined. Then these networks are used to apply the proposed measure and verify its efficiency. The efficiency of the measure is analyzed by the damage caused to the network through three measures: attribute load, robustness and toughness.

4.4.1 Targeting Method and Network Damage

Five traditional measures of centrality were selected : degree centrality (DC), closeness centrality (CC), betweenness centrality (BC), pagerank (PR), eigenvector centrality (EC); Euclidean distance method and the measure of Network Capital (NC) to verify the efficiency of the proposed measure in terms of the damage done to the network. Each of these seven measurements was modeled by a targeting method that starts with the entire network. In order to evaluate the deterioration in the network.

The measures of centrality are calculated for all vertices in the initial network, and the vertex with highest centrality measure is removed. For the resulting network, the centrality measures of all vertices is recalculated and again remove the highest ranked and so forth, until the required number q of vertices has been removed. This process is known as sequential targeted attack [90]. The degree centrality and betweenness centrality used in the Euclidean distance method were obtained by the sequential process. The ED_t and NC were also obtained for initial network G and also using the sequential targeted attack, the node that most reduced the network energy or the network capital was eliminated from the network.

In unweighted networks, damage to the network is generally studied, considering increasingly larger fractions of nodes in the network G being removed. This removal is chosen following different strategies [78, 90]. In particular, the topological integrity of the network n_q/n has been studied, where n_q is the size of the largest component after removing q nodes and n is the size of the original network, $n = ||V||$. A measure that assesses the robustness of the network, that is, the ability to resist attacks, is defined by Iyer et al. [90]. For a given sequence of target nodes removal, Iyer et al. [90] defined robustness considering the mean topological integrity of the network as the number of removed nodes range from 1 to n. For a given sequence of nodes' removal, here this measure is extended by considering the robustness as a function of the maximum number of nodes removed in an attack, q_{max}, as follows:

$$R(q_{max}) = \frac{1}{q_{max}} \sum_{q=1}^{q_{max}} \frac{n_q}{n}. \tag{33}$$

Since the robustness does not take into consideration neither the number of links nor their strengths and, it is considered that this measure is not suitable for weighted networks. Dall'Asta et al. [78] also agrees that in weighted networks the damage quantification should consider the presence of the edge weights, furthermore, for the authors, the strength of the connections is probably an important indicator of the network's functionality.

To include the importance and toughness of relationships, attributed by the weights of the connections [95] and inspired by the robustness metric, it is proposed a new damage measure, called toughness, that includes the strength of the connections. This metric also works for unweighted, where all edges' weights are considered to be equal to 1. In the toughness metric, It is not looked specifically at the largest component after removing q nodes, but at the component that has the largest sum of edges' weights, w_q, after such removal. With this perspective, it is assumed that the functionality of the network is dependent on the strength of the connections, w_q. For this reason, for a given sequence of nodes' removal, the toughness of a network to this attack is defined by:

$$R_w(q_{max}) = \frac{1}{q_{max}} \sum_{q=1}^{q_{max}} \frac{w_q}{w} \tag{34}$$

w is the sum of all edges weights of G.

A measure similar to robustness was also developed to analyze the loss of network resources. Since resources (attributes) as well as connections are important for the functioning of the network, i.e. both have particular value to the overall operation of the network. this measure is called of attributes load, A, and, for a given sequence of nodes' removal, it is defined as follows:

$$A(q_{max}) = \frac{1}{q_{max}} \sum_{q=1}^{q_{max}} \frac{k_q}{k} \tag{35}$$

k_q is the largest sum of attribute presented in a component after removing q nodes and k is the sum of nodes' attributes in G.

In the following subsections it is investigated the effect in the three network damage measures of removing vertices according to specified targeting methods based on classical centrality measures, NC, Euclidean distance method and on the energy disruptive centrality measure. In the present figures, matrices are used to represent the loss of attributes and the disruption of the network when q nodes are removed, according to some targeting criteria. The metrics presented in each row of the matrices are ordered conforming to the attribute load, robustness or toughness of the network, where q_{max} varies between 1 and the total number of nodes. It is emphasized that the most efficient targeting metrics are the ones that obtain the lower value for a given damage measure.

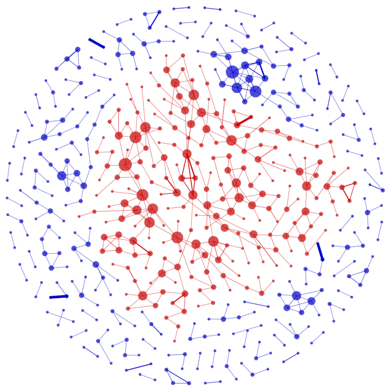
4.4.2 Network of criminals encounters

Data for this work were obtained from the electronic anklet movement over a period of thirty days in a major city in the Northeast of Brazil. Through position, in latitude and longitude, over time, it is built a network of encounters between convicts wearing electronic anklets. It is assumed that an encounter occurs when two electronic anklets remain at a given distance (10 meters or less) from each other for a certain period of time (5 minutes or more). The meetings that took place at the convict's residence, meetings between married people and meetings that took place in places that the convicts must attend by judicial decision were not considered in the formation of the network.

The network formed is composed of 523 criminals wearing electronic anklets and 539 undirected connections. It contains 96 connected components (the principal component has 216 nodes). The network is far from complete, the density is only 0.0039, the principal component is denser $(D = 0,0125)$. The clustering coefficient measures how dense is a node's neighborhood, indicating the ability a node has to connect its neighbors. The network and its principal component present an average clustering of 0.1498 and 0.1993, respectively, indicating relevant neighborhood insertion. In this work, it is considered both unweighted and weighted networks. The weighting of networks occurs in two ways, the first the strength of the connections indicates the number of days the ankle monitors met and the second, how long, in hours, they spent together. The weights, days and hours, were normalized by having all of them divided by the largest weight in the network.

For the nodes' attributes, that measures features of the criminals in the network, it is used information on displacements: maximum speed, average speed, total distance and surrounding area. In order to group the 4 measurements and define a single weight for the nodes, it is used factor analysis. Factor analysis is a technique that aims to reduce the dimensionality of a data set [99]. Brantingham and Brantingham [84] comment that criminals tend to act within their awareness space, paths and places that are part of their routines. Chaturapruek et al. [85] mention that some criminal displacement data indicates that they are willing to travel greater distances in search of more attractive targets. This work starts from the premise that those who have a greater possibility of commuting are more likely to commit crimes, they present a greater threat. Factor analysis assigned the weight 0.592 for the surrounding area, 0.646 for maximum speed, 0.927 for the sum of the distance and 0.928 for average speed. The resulting node attribute was also normalized, by having all of them divided by the maximum attribute in the network.

Figure 15 – The full network visualization with layout Fruchterman-Reingold. The main component is highlighted in red, the radius of the nodes is proportional to their degree and the thickness of the edges is proportional to the frequency of encounters in hours.



Source: Andrade et al. [73].

Even though the information on the displacement of the actors constitute one important factor in the detection of crime, any other resources network actors owned, appropriately used for law enforcement, can be used in weighting the nodes. It is used ankle monitors displacement data because this was the only relevant information available. Bright et al. [15] and Bright et al. [100] defined the weights of the attributes as amounts of resources that the authors of the network have. In the drug trafficking network, for example, Bright et al. [15] classified resources as tangible: money, drugs, premises, equipment, precursors; and intangible: information, skill/knowledge, labor.

The degree distribution and the weighted (by day or hours) degree distribution of the network shown in Figure 16 present power law distributions. It can be seen by the node radius shown in Figure 15 that only a few nodes are notable. This heterogeneous characteristic, presented by networks of free scales, makes the network resistant to random attacks and fragile to targeted attacks [101–104].

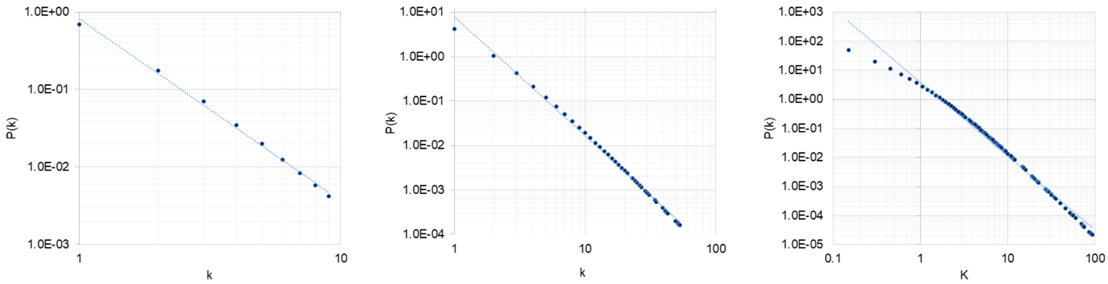
The effect of removing of the nodes on the network of criminals encounters will be analyzed in three ways: (1) without considering the weight of the edges, unweighted network (UW); (2) weighted based on the number of days of encounters (WD); and (3) weighted based on the number of hours of encounters (WH).

4.4.2.1 Attribute load

Analyzing the loss of nodes' attributes after removing some sequence of nodes is a way to check for network functionality damage, since they contribute to the full operation of the

Figure 16 – Degree distributions.

(a) Unweighted (p-value = 0.00) (b) Weighted by days (p-value = 1.22E-192) (c) weighted by hours (p-value = 3.40E-268)



Source: Andrade et al. [73]

network. The network of ankle monitors is composed of several components, each of which has an attribute load, that corresponds to the sum of the individual attribute weights. Our analysis of the loss of attribute load is directed at the component that has the highest load after removing a certain number of nodes, following Equation 35.

Figure 17 shows how each targeting method under analysis causes the loss of the attribute after the removal of q_{max} nodes. The measures are listed in increasing order of $A(n)$, which is the average relative loss of attribute load when all nodes are removed according to the given targeting method and is displayed in the right side of the figure. The smaller this average, the greater the efficiency of the measure in removing first the most important nodes in the network. It is possible to note that among the three most efficient measures in the three cases: UW, WD and WH; two are: ED_5 and ED_2. The third measure among the most efficient is BC. Figure 17 also shows that the unweighted case is more sensitive to attack. The loss of attribute load occurs on average, considering all target methods, 2.2 times more in the unweighted network than in the day-weighted network and 2.6 times more than in the hour-weighted network.

Removing a single node on the network does not, by itself, significantly change a component's attribute load. This is particularly true for components whose attribute load is much greater than the attribute weight of a single node. However, when removing it, other nodes connected to the component through it disconnect, causing a greater impact on the attribute load. Therefore, the most efficient measures, ED_5 and ED_2, are able to disconnect from the component with the highest attribute load, nodes that cause the greatest reduction in that load. As the attributes are related to the ankle monitors displacement capacity, a lower attribute load in the component means less potentiality for criminal actions, given the hypothesis that the crime occurs along the convict's trajectory or is facilitated by its easiness of displacement.

Figure 18 suggests that there is no correlation between the node's attribute weight and the targeted removal of nodes according to traditional measures of centrality. Thus, the node centrality is not related to the node's attribute weight. This result reaffirms the discussion presented in Section 4.2, that individuals with high cognitive load or attribute weight are not necessarily those with greater centrality. In contrast, there is a greater propensity for the first target attacks to be nodes with a high attribute weight, this is particularly true for ED with lower t and considering the weighting of the edges. For t high, it is observed that the nodes that stand out the most, having high attribute weight, tend to be removed later.

The efficiency of the BC measure in decreasing the attribute load is associated with its power to dismantle the network. In contrast, it can be noticed that the targeted removal of nodes according to the measures derived from the ED_t have a certain correlation with the node's attribute weight. This correlation is even greater in weighted networks. In these networks, the average of the distances of the shortest paths are greater, consequently, the interaction force between the nodes is lower, especially when t is less than 1. Therefore, NE_t is constituted mainly by the sum of the attribute weights, as shown in Equation 30.

Figure 17 – Attribute load as a function of the total number of nodes removed q_{max}.

(a) Unweighted case

k_q/k \ q_{max}	1	2	3	4	5	6	7	8	9	10	20	50	100	200	500	A(n)
ED_1	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.03	0.01	0.01	0.00	0.00	0.00669
ED_5	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.04	0.01	0.01	0.00	0.00	0.00677
ED_2	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.04	0.01	0.01	0.00	0.00	0.00679
NC(RSL=0.99)	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.04	0.01	0.01	0.00	0.00	0.00692
NC(RSL=0.5)	0.26	0.26	0.19	0.11	0.11	0.06	0.05	0.05	0.05	0.05	0.04	0.01	0.01	0.00	0.00	0.00728
$ED_{1/2}$	0.26	0.15	0.11	0.11	0.11	0.08	0.05	0.05	0.05	0.05	0.04	0.02	0.01	0.00	0.00	0.00733
CC	0.20	0.20	0.10	0.10	0.09	0.06	0.06	0.06	0.06	0.04	0.03	0.02	0.01	0.01	0.00	0.00876
BC	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.03	0.01	0.01	0.01	0.00	0.00990
$ED_{1/5}$	0.32	0.28	0.28	0.25	0.18	0.16	0.15	0.15	0.15	0.12	0.06	0.02	0.01	0.00	0.00	0.01166
NC(RSL=0.1)	0.32	0.28	0.21	0.19	0.16	0.16	0.15	0.15	0.15	0.15	0.09	0.02	0.01	0.00	0.00	0.01235
DC	0.42	0.26	0.26	0.25	0.25	0.24	0.22	0.22	0.22	0.22	0.07	0.04	0.02	0.01	0.00	0.01429
EC	0.42	0.28	0.27	0.27	0.27	0.27	0.27	0.27	0.18	0.18	0.07	0.02	0.01	0.00	0.00	0.01430
PR	0.26	0.26	0.24	0.23	0.21	0.21	0.21	0.21	0.21	0.21	0.05	0.03	0.03	0.01	0.00	0.01520
Euclidean A	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.18	0.18	0.07	0.04	0.02	0.01	0.00	0.01589
Euclidean B	0.42	0.42	0.41	0.20	0.20	0.20	0.20	0.20	0.20	0.19	0.18	0.17	0.13	0.08	0.01	0.07913

(b) Weighted by day

k_q/k \ q_{max}	1	2	3	4	5	6	7	8	9	10	20	50	100	200	500	A(n)
ED_5	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.04	0.02	0.01	0.00	0.00	0.00682
ED_2	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.04	0.01	0.01	0.00	0.00	0.00729
BC	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.03	0.01	0.01	0.01	0.00	0.00995
ED_1	0.27	0.15	0.11	0.11	0.11	0.06	0.05	0.05	0.05	0.05	0.04	0.03	0.01	0.01	0.00	0.01066
CC	0.20	0.20	0.20	0.10	0.10	0.10	0.10	0.10	0.09	0.06	0.04	0.03	0.01	0.01	0.00	0.01179
PR	0.26	0.26	0.15	0.15	0.14	0.14	0.11	0.11	0.11	0.11	0.05	0.03	0.03	0.01	0.00	0.01364
DC	0.27	0.27	0.27	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.13	0.05	0.03	0.01	0.00	0.02277
NC(RSL=0.99)	0.41	0.41	0.39	0.39	0.39	0.30	0.29	0.29	0.29	0.29	0.11	0.06	0.03	0.01	0.00	0.02294
NC(RSL=0.5)	0.41	0.41	0.41	0.41	0.39	0.30	0.30	0.29	0.29	0.29	0.13	0.07	0.03	0.01	0.00	0.02838
EC	0.42	0.41	0.40	0.27	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.11	0.02	0.01	0.00	0.03092
Euclidean A	0.42	0.27	0.27	0.27	0.27	0.27	0.25	0.25	0.25	0.25	0.25	0.08	0.04	0.02	0.00	0.03503
$ED_{1/2}$	0.41	0.41	0.41	0.41	0.40	0.40	0.40	0.40	0.39	0.30	0.28	0.13	0.08	0.04	0.00	0.04755
NC(RSL=0.1)	0.42	0.41	0.41	0.41	0.41	0.41	0.40	0.40	0.40	0.40	0.38	0.13	0.08	0.04	0.00	0.05298
$ED_{1/5}$	0.42	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.40	0.38	0.38	0.16	0.11	0.04	0.00	0.06168
Euclidean B	0.42	0.42	0.41	0.41	0.41	0.40	0.40	0.40	0.40	0.40	0.37	0.34	0.28	0.19	0.01	0.13826

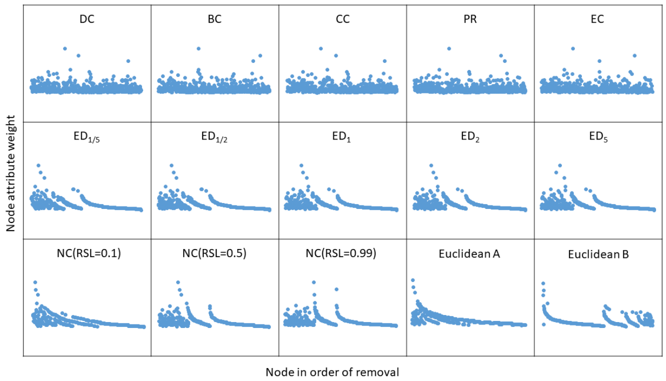
(c) Weighted by hour

k_q/k \ q_{max}	1	2	3	4	5	6	7	8	9	10	20	50	100	200	500	A(n)
ED_5	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.03	0.01	0.01	0.00	0.00	0.00698
ED_2	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.05	0.04	0.02	0.01	0.01	0.00	0.00925
BC	0.27	0.15	0.15	0.11	0.09	0.08	0.05	0.05	0.05	0.05	0.04	0.01	0.01	0.01	0.00	0.00988
PR	0.42	0.22	0.22	0.14	0.14	0.14	0.14	0.13	0.11	0.11	0.06	0.04	0.02	0.01	0.00	0.01442
CC	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.09	0.04	0.02	0.01	0.00	0.02145
ED_1	0.27	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.13	0.13	0.08	0.05	0.04	0.02	0.00	0.02211
DC	0.42	0.42	0.42	0.32	0.32	0.32	0.32	0.19	0.19	0.19	0.18	0.11	0.04	0.01	0.00	0.02887
NC(RSL=0.99)	0.42	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.17	0.05	0.04	0.01	0.00	0.03227
EC	0.42	0.42	0.42	0.40	0.32	0.32	0.32	0.32	0.32	0.19	0.18	0.11	0.04	0.01	0.00	0.03228
Euclidean A	0.42	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.26	0.08	0.03	0.02	0.00	0.03292
NC(RSL=0.5)	0.42	0.42	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.30	0.08	0.07	0.02	0.00	0.04136
Euclidean B	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.24	0.13	0.07	0.03	0.01	0.05155
NC(RSL=0.1)	0.42	0.41	0.41	0.41	0.41	0.40	0.40	0.40	0.40	0.40	0.38	0.15	0.10	0.04	0.00	0.05880
$ED_{1/2}$	0.42	0.42	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.40	0.40	0.15	0.10	0.04	0.00	0.05980
$ED_{1/5}$	0.42	0.42	0.41	0.41	0.41	0.41	0.40	0.40	0.40	0.40	0.38	0.15	0.12	0.04	0.00	0.06127

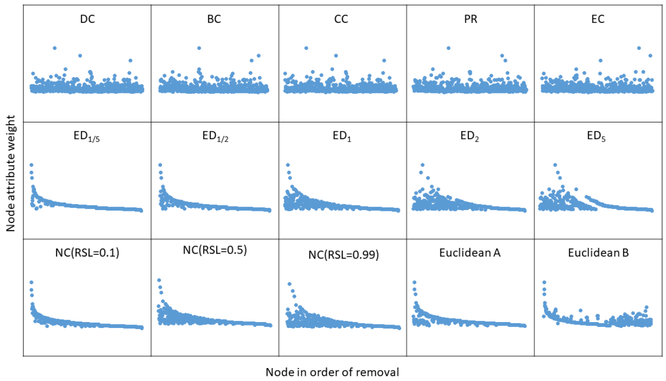
Source: Andrade et al. [73].

Figure 18 – Scatterplot of node removed and node attribute weighting.

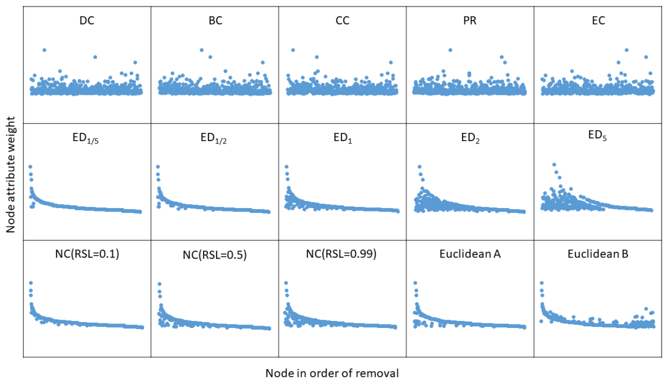
(a) Unweighted case



(b) Weighted by day



(c) Weighted by hour



Source: Andrade et al. [73].

4.4.2.2 Robustness

The robustness analyzes the damage to the network by the loss of nodes in the largest component. This is the most commonly used form of damage analysis. In Figure 19, it is noted that among the three most efficient measures, in the three cases, UW, WD and WH, two are: ED_5 and ED_2. BC is also among the top three most efficient, but only in weighted cases. It is important highlighted that, the best attack strategy was not entirely a topological intervention, but a measure that mixes characteristics of the network topology with nodes' attributes, as the one proposed in this work. Surpassing the BC, cited in several articles as the best strategy to target the removal of nodes in order to maximize disruption of networks [90, 100].

Figure 19 – Robustness as a function of the total number of nodes removed q_{max}.

(a) Unweighted case

n_q/n \ q_{max}	1	2	3	4	5	6	7	8	9	10	20	50	100	200	500	R(n)
ED_2	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.03	0.02	0.01	0.00	0.00	0.00690
NC(RSL=0.99)	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.04	0.01	0.01	0.00	0.00	0.00694
ED_5	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.04	0.02	0.01	0.00	0.00	0.00701
CC	0.20	0.20	0.10	0.10	0.09	0.06	0.06	0.06	0.06	0.05	0.03	0.02	0.01	0.00	0.00	0.00708
NC(RSL=0.5)	0.25	0.25	0.19	0.11	0.11	0.06	0.05	0.05	0.05	0.05	0.04	0.02	0.01	0.00	0.00	0.00739
$ED_{1/2}$	0.25	0.15	0.11	0.11	0.11	0.08	0.05	0.05	0.05	0.05	0.04	0.02	0.01	0.00	0.00	0.00760
ED_1	0.25	0.15	0.11	0.11	0.11	0.08	0.05	0.05	0.05	0.05	0.04	0.02	0.01	0.00	0.00	0.00760
BC	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.05	0.03	0.01	0.01	0.01	0.00	0.00856
$ED_{1/5}$	0.32	0.27	0.27	0.24	0.18	0.16	0.16	0.16	0.16	0.13	0.06	0.02	0.01	0.00	0.00	0.01197
EC	0.41	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.18	0.18	0.07	0.02	0.01	0.00	0.00	0.01259
NC(RSL=0.1)	0.32	0.27	0.21	0.19	0.16	0.16	0.16	0.16	0.16	0.16	0.10	0.02	0.02	0.00	0.00	0.01283
PR	0.25	0.25	0.23	0.23	0.21	0.21	0.21	0.20	0.20	0.20	0.05	0.03	0.02	0.00	0.00	0.01347
DC	0.41	0.25	0.25	0.25	0.25	0.24	0.22	0.22	0.22	0.22	0.07	0.04	0.02	0.00	0.00	0.01379
Euclidean A	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.18	0.17	0.07	0.04	0.02	0.01	0.00	0.01632
Euclidean B	0.41	0.41	0.41	0.20	0.20	0.20	0.20	0.20	0.20	0.19	0.18	0.17	0.13	0.08	0.01	0.07896

(b) Weighted by day

n_q/n \ q_{max}	1	2	3	4	5	6	7	8	9	10	20	50	100	200	500	R(n)
ED_5	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.03	0.01	0.01	0.00	0.00	0.00707
ED_2	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.03	0.01	0.01	0.00	0.00	0.00756
BC	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.05	0.03	0.01	0.01	0.01	0.00	0.00815
CC	0.20	0.20	0.20	0.10	0.10	0.10	0.10	0.10	0.09	0.06	0.04	0.03	0.01	0.00	0.00	0.01014
ED_1	0.27	0.15	0.11	0.11	0.11	0.06	0.05	0.05	0.05	0.05	0.04	0.03	0.02	0.01	0.00	0.01125
PR	0.25	0.25	0.15	0.15	0.14	0.14	0.11	0.11	0.11	0.11	0.05	0.03	0.02	0.00	0.00	0.01202
DC	0.27	0.27	0.27	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.12	0.05	0.03	0.00	0.00	0.02114
NC(RSL=0.99)	0.41	0.41	0.39	0.39	0.39	0.30	0.29	0.29	0.29	0.29	0.11	0.06	0.04	0.01	0.00	0.02373
EC	0.41	0.41	0.39	0.27	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.11	0.02	0.00	0.00	0.02870
NC(RSL=0.5)	0.41	0.41	0.41	0.41	0.39	0.30	0.30	0.29	0.29	0.29	0.14	0.07	0.03	0.01	0.00	0.02961
Euclidean A	0.41	0.27	0.27	0.27	0.27	0.27	0.25	0.25	0.25	0.25	0.25	0.08	0.04	0.02	0.00	0.03609
$ED_{1/2}$	0.41	0.41	0.41	0.41	0.40	0.40	0.40	0.40	0.39	0.30	0.16	0.13	0.08	0.04	0.00	0.05010
NC(RSL=0.1)	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.40	0.40	0.40	0.38	0.14	0.09	0.04	0.00	0.05570
$ED_{1/5}$	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.40	0.38	0.38	0.16	0.11	0.04	0.00	0.06444
Euclidean B	0.41	0.41	0.41	0.41	0.41	0.4	0.4	0.4	0.4	0.4	0.37	0.34	0.29	0.20	0.01	0.14178

(c) Weighted by hour

n_q/n \ q_{max}	1	2	3	4	5	6	7	8	9	10	20	50	100	200	500	R(n)
ED_5	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.04	0.03	0.02	0.01	0.00	0.00	0.00724
BC	0.27	0.15	0.14	0.11	0.09	0.08	0.05	0.05	0.05	0.05	0.03	0.01	0.01	0.01	0.00	0.00836
ED_2	0.20	0.20	0.10	0.10	0.09	0.06	0.05	0.05	0.05	0.05	0.04	0.02	0.01	0.01	0.00	0.00974
PR	0.41	0.22	0.22	0.14	0.14	0.14	0.14	0.13	0.11	0.11	0.06	0.04	0.01	0.00	0.00	0.01270
CC	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.06	0.03	0.02	0.01	0.00	0.01995
ED_1	0.27	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.13	0.13	0.05	0.05	0.04	0.03	0.00	0.02339
DC	0.41	0.41	0.41	0.32	0.32	0.32	0.32	0.19	0.19	0.19	0.17	0.11	0.04	0.00	0.00	0.02708
EC	0.41	0.41	0.41	0.39	0.32	0.32	0.32	0.31	0.31	0.19	0.18	0.11	0.04	0.01	0.00	0.03007
NC(RSL=0.99)	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.17	0.05	0.04	0.02	0.00	0.03311
Euclidean A	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.31	0.09	0.07	0.03	0.00	0.03415
NC(RSL=0.5)	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.31	0.09	0.07	0.03	0.00	0.04327
Euclidean B	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.17	0.05	0.04	0.02	0.00	0.05329
NC(RSL=0.1)	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.40	0.40	0.40	0.38	0.16	0.11	0.04	0.00	0.06164
$ED_{1/2}$	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.40	0.38	0.16	0.11	0.04	0.00	0.06260
$ED_{1/5}$	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.40	0.40	0.38	0.16	0.12	0.04	0.00	0.06409

Source: Andrade et al. [73].

The measure also proved to be more efficient than the network capital construct as described by Schwartz and Rouselle [16], which also proposed to attack criminal networks, incorporating weighting of actors and network connections. ED_2 and ED_5 are the measures with the greatest power to disrupt the network, reduce the largest component. In other words, these measures are able to interrupt the flow of communication between those being monitored, making it difficult to plan criminal actions. Since ED_2 and ED_5 were also efficient in reducing the network node's ability to move around, it is expected that due to the weakening of communications and restricted displacement, the potential of crime will decrease.

It can be also shown in this analysis that the unweighted case is more sensitive to attack. The loss of nodes in the largest component occurs on average, considering all targeting methods, 2.3 times more in the unweighted network than in the day-weighted network and 2.7 times more than in the hour-weighted network.

4.4.2.3 Toughness

It is known that, when removing nodes from the network, the edges that connected to these nodes are also removed. As the connections are related to the interactions between the nodes, the stronger the relationship, the more important and resilient it is. As an important indicator of network functionality, losing connections means loss of interactions, leading to reduced network functionality, i.e., damage. It is called toughness the loss of edges (weighted or not) of the component with the largest sum of edges.

Figure 20 shows the network toughness for the three cases studied: UW, WD and WH. In the weighted case, as in the previous analyses, the three measures ED_5, ED_2 and BC remain as the most efficient targeting methods. In the unweighted case, three new measures took the position of more efficient CC, $ED_{1/2}$ and ED_1. Thus, although ED_5, ED_2 and BC remove nodes faster from the largest component, the number of edges lost after removing the nodes is less than the number of edges lost after removing nodes by the less efficient measures in terms of robustness: CC, $ED_{1/2}$ and ED_1.

In the unweighted network, where each connection has the same importance, the CC, $ED_{1/2}$ and ED_1 measures were the most efficient since they removed a greater number of connections, however, it is not known whether those connections are the ones that most contribute to the functioning of the network. Thus, the lowest toughness occurs in a quantitative manner, due to the greater removal of connections and not in a qualitative manner, removal of the most important edges. In contrast, in the weighted networks, the lowest toughness occurs by removing a number of connections that are more intense or strong. In this case, ED_5 is observed as the most efficient measure.

Finally, it is point out that the toughness of the unweighted network is on average, considering all targeting methods, 2.7 times less than the toughness of the day-weighted network and 3.4 times less than the toughness of the hour-weighted network.

Figure 20 – Toughness as a function of the total number of nodes removed q_{max}.

(a) Unweighted case

W_q/W \ q_{max}	1	2	3	4	5	6	7	8	9	10	20	50	100	200	500	$R_w(n)$
CC	0.27	0.25	0.14	0.12	0.11	0.08	0.08	0.07	0.07	0.06	0.03	0.01	0.01	0.00	0.00	0.00611
$ED_{1/2}$	0.32	0.20	0.14	0.14	0.14	0.10	0.08	0.08	0.08	0.08	0.05	0.02	0.01	0.00	0.00	0.00722
ED_1	0.27	0.25	0.14	0.14	0.11	0.08	0.08	0.08	0.08	0.06	0.04	0.02	0.01	0.00	0.00	0.00752
NC(RSL=0.5)	0.32	0.32	0.24	0.14	0.14	0.08	0.08	0.08	0.08	0.08	0.06	0.02	0.01	0.00	0.00	0.00757
NC(RSL=0.99)	0.27	0.27	0.14	0.14	0.11	0.08	0.08	0.08	0.08	0.06	0.05	0.02	0.01	0.00	0.00	0.00791
ED_5	0.27	0.27	0.14	0.14	0.11	0.08	0.08	0.08	0.08	0.06	0.05	0.02	0.01	0.00	0.00	0.00793
ED_2	0.27	0.27	0.14	0.14	0.11	0.08	0.08	0.08	0.08	0.06	0.05	0.02	0.01	0.00	0.00	0.00801
BC	0.27	0.27	0.14	0.14	0.11	0.08	0.08	0.08	0.08	0.06	0.04	0.02	0.01	0.01	0.00	0.00952
$ED_{1/5}$	0.41	0.34	0.34	0.31	0.23	0.21	0.19	0.19	0.19	0.14	0.06	0.03	0.01	0.00	0.00	0.01156
EC	0.54	0.34	0.33	0.33	0.31	0.31	0.31	0.31	0.20	0.20	0.07	0.02	0.00	0.00	0.00	0.01212
NC(RSL=0.1)	0.41	0.34	0.27	0.24	0.21	0.21	0.19	0.19	0.19	0.19	0.11	0.02	0.01	0.00	0.00	0.01258
DC	0.54	0.32	0.32	0.30	0.30	0.28	0.25	0.25	0.25	0.25	0.07	0.04	0.01	0.00	0.00	0.01325
PR	0.32	0.32	0.29	0.28	0.26	0.26	0.26	0.24	0.24	0.24	0.06	0.04	0.02	0.00	0.00	0.01432
Euclidean A	0.27	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.22	0.21	0.07	0.04	0.02	0.01	0.00	0.01599
Euclidean B	0.54	0.54	0.54	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.25	0.24	0.17	0.10	0.01	0.10263

(b) Weighted by day

W_q/W \ q_{max}	1	2	3	4	5	6	7	8	9	10	20	50	100	200	500	$R_w(n)$
ED_5	0.29	0.29	0.15	0.15	0.12	0.07	0.07	0.07	0.07	0.05	0.05	0.01	0.01	0.00	0.00	0.00739
ED_2	0.29	0.29	0.15	0.14	0.12	0.07	0.07	0.07	0.07	0.05	0.05	0.02	0.01	0.00	0.00	0.00850
BC	0.29	0.29	0.15	0.15	0.12	0.07	0.07	0.07	0.07	0.06	0.04	0.01	0.01	0.01	0.00	0.00919
CC	0.29	0.26	0.26	0.15	0.14	0.14	0.14	0.14	0.12	0.07	0.06	0.03	0.02	0.00	0.00	0.01121
ED_1	0.38	0.21	0.15	0.15	0.15	0.07	0.07	0.07	0.07	0.07	0.05	0.04	0.02	0.01	0.00	0.01312
PR	0.34	0.33	0.21	0.21	0.20	0.20	0.15	0.15	0.15	0.15	0.07	0.04	0.03	0.00	0.00	0.01333
DC	0.38	0.38	0.38	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.18	0.06	0.04	0.00	0.00	0.02718
NC(RSL=0.99)	0.58	0.57	0.55	0.55	0.55	0.42	0.41	0.41	0.41	0.41	0.15	0.08	0.05	0.01	0.00	0.02996
NC(RSL=0.5)	0.58	0.58	0.57	0.57	0.55	0.42	0.42	0.41	0.41	0.41	0.19	0.10	0.04	0.01	0.00	0.03758
EC	0.58	0.57	0.56	0.38	0.36	0.36	0.36	0.36	0.36	0.36	0.35	0.15	0.03	0.00	0.00	0.03809
Euclidean A	0.38	0.38	0.38	0.38	0.38	0.36	0.36	0.36	0.36	0.36	0.36	0.11	0.05	0.02	0.00	0.04491
$ED_{1/5}$	0.42	0.41	0.41	0.41	0.41	0.41	0.41	0.40	0.40	0.38	0.38	0.16	0.11	0.04	0.00	0.06168
$ED_{1/2}$	0.58	0.58	0.58	0.57	0.57	0.57	0.57	0.57	0.55	0.42	0.41	0.18	0.11	0.05	0.00	0.06429
NC(RSL=0.1)	0.58	0.58	0.58	0.58	0.57	0.57	0.57	0.56	0.56	0.56	0.54	0.18	0.12	0.05	0.00	0.07110
Euclidean B	0.58	0.58	0.58	0.57	0.56	0.56	0.56	0.56	0.56	0.56	0.54	0.47	0.40	0.25	0.00	0.18731

(c) Weighted by hour

W_q/W \ q_{max}	1	2	3	4	5	6	7	8	9	10	20	50	100	200	500	$R_w(n)$
ED_5	0.28	0.28	0.15	0.15	0.12	0.08	0.07	0.07	0.07	0.07	0.05	0.02	0.01	0.01	0.00	0.00932
BC	0.38	0.20	0.19	0.15	0.12	0.12	0.07	0.07	0.07	0.07	0.07	0.02	0.01	0.01	0.01	0.01267
ED_2	0.28	0.26	0.15	0.15	0.12	0.08	0.07	0.07	0.07	0.07	0.06	0.04	0.02	0.01	0.00	0.01334
PR	0.57	0.29	0.29	0.20	0.20	0.20	0.20	0.18	0.15	0.15	0.07	0.05	0.02	0.00	0.00	0.01513
CC	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.12	0.05	0.02	0.01	0.00	0.02660
ED_1	0.38	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.18	0.18	0.11	0.07	0.04	0.03	0.00	0.02929
DC	0.57	0.57	0.57	0.44	0.44	0.44	0.44	0.28	0.28	0.28	0.25	0.14	0.05	0.01	0.00	0.03670
EC	0.57	0.57	0.57	0.55	0.44	0.44	0.44	0.44	0.44	0.28	0.26	0.14	0.06	0.01	0.00	0.04117
Euclidean A	0.57	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.37	0.10	0.05	0.03	0.00	0.04386
NC(RSL=0.99)	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.25	0.07	0.04	0.02	0.00	0.04433
NC(RSL=0.5)	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.43	0.10	0.09	0.03	0.00	0.05694
Euclidean B	0.38	0.38	0.38	0.38	0.38	0.38	0.37	0.37	0.37	0.37	0.34	0.17	0.10	0.05	0.00	0.06939
NC(RSL=0.1)	0.57	0.57	0.57	0.57	0.57	0.56	0.56	0.56	0.56	0.56	0.53	0.22	0.13	0.04	0.00	0.07718
$ED_{1/2}$	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.56	0.56	0.56	0.22	0.13	0.04	0.00	0.07936
$ED_{1/5}$	0.57	0.57	0.57	0.57	0.57	0.57	0.56	0.56	0.56	0.56	0.54	0.22	0.15	0.04	0.00	0.08201

Source: Andrade et al. [73].

4.4.2.4 Relationship between violated crimes and targeting methods

Table 21 presents the crimes committed by the first 20 criminals removed for each of the targeting methods, considering recidivism. The violated laws[1] are: (A) 2848; (B) 8069; (C) 12850; (D) 10826; (E) 11340; (F) 11343; (G) 10671; (H) 1521; (I) 9503; and (J) 8176. Each column represents an Art. of that law. It is observed a greater concentration of infractions in Art. 157 (A) of the Penal Code and in Arts. 33 and 35 (F) of the Narcotic Law. This concentration occurs mainly in the network when it is considered the unweighted network. Art. 157 of the penal code refers to subtraction of other people's movable assets through a serious threat. Art. 33 of the narcotics law refers to manipulation, sale or offer of drugs, while Art. 35 refers to association of two or more people for the purpose of practicing, repeatedly or not, any of the crimes provided for in Art. 33.

In the last column of the table, the expected average sentencing time (AST) is shown, in descending order, for the twenty convicts removed from the network by the corresponding measure in the first column.

In the unweighted case, DC, $NC(RSL = 0.1)$ and ED_2 are the measures that removed nodes with higher AST. Convicts who violated Art. 33 of the Narcotic Law have high DC, i.e., several connections, demonstrating a great capacity for drug distribution, if the convict still practices this crime. Additionally, given that the removal of nodes following DC with an attack strategy is not as efficient in reducing the attributes, which are related to the displacement of ankle monitors, it suggests that the distribution of drugs, if it still occurs through these convicts, happens at specific points. $NC(RSL = 0.1)$ and ED_2 removed convicts who acted primarily in armed robbery, Art. 157 of the penal code. These two measures are heavily influenced by attributes. However, ED_2 when used as a targeting method has a greater capacity to remove nodes that are fundamental to the structure of the network. As shown in the Sections 4.4.2.1 and 4.4.2.2, the removal of nodes according to ED_2 caused a greater loss of attributes and reduction of network robustness.

In the case of the network weighted by days of encounters or hours of encounters, ED_5 and BC are the measures that removed those convicted with higher AST. Anklet monitors removed by these measures were involved mainly in armed robbery, Art. 157 of the penal code. ED_5 also stands out for removing convicts for culpable reception, Art. 180 of the penal code,

[1] The aforementioned laws can be found on the Legislation Portal of the Federal Government of Brazil. Link: http://www4.planalto.gov.br/legislacao/

and BC significantly removes convicts for drug trafficking, Art. 33 of the Narcotic Law. As shown in the analysis of the Subsections 4.4.2.1, 4.4.2.2 and 4.4.2.3, BC and ED_5 are efficient in disrupting the network, so attacking the network following these measures can reduce or create barriers to future recurrences of crimes.

The number of infractions of Art. 157 of the penal code is significant among the twenty nodes removed in all measures used as a method of attack, mainly in the unweighted network. There are 171 convicted in the network for violating this article. It is removed from the network the twenty convicts with the highest recurrence and verified the damage caused and the efficiency of this strategy. In the unweighted and weighted network, the largest attribute load after the twenty removals was 0.25. By the most efficient measures ED_2 and ED_5, with the first node removed, the largest attribute load was 0.20 for both. The robustness of the weighted and unweighted network was 0.25, i.e., the largest component has only 25% of all nodes in the network before removing the twenty nodes. Again, measures, ED_2 and ED_5, cause greater damage only with the first node removed, as shown in the Figure 19. In relation to toughness, after removing the twenty convicts with greater recidivism in Art. 157 the toughness of the unweighted network was 0.31 and in the other two weighted networks, it was 0.33. Again this strategy proves inefficient. The toughness of the unweighted network after removing the first node according to the measures, ED_2 and ED_5, was 0.27, in the day-weighted network, 0.29, and in the hour-weighted network, 0.28.

Regarding the length of the sentence, it can be noted that the 20 convicts removed from the unweighted network spent on average considering all targeting methods 6.7 years longer in prison compared to the 20 removed from the day-weighted network and 7.1 more years compared to the 20 removed from the hour-weighted network.

Figure 21 – Crimes committed by the 20 ankle monitors removed from the network.

(a) Unweighted edges

	A																								B	C	D			E		F				G	H	I	J	
	29	80	121	129	137	147	155	157	163	171	176	180	244	288	299	304	307	311	329	330	331	349	352	354	244	2	12	14	16	7	5	28	33	35	40	41	4	309	1	AST
DC	0	0	5	4	0	0	3	22	5	1	0	2	0	2	1	0	2	0	3	4	6	1	1	0	4	1	1	1	8	0	0	2	23	9	2	0	0	0	0	31.6
NC(RSL=0.1)	0	0	5	3	1	2	3	32	1	1	0	5	1	2	3	0	2	0	5	4	6	1	0	0	2	1	1	3	6	0	4	1	15	6	2	0	0	5	1	31.0
ED_2	0	0	5	4	1	2	2	31	1	1	0	6	0	3	1	0	2	0	3	4	4	1	0	0	3	1	1	2	7	5	1	1	16	6	2	0	0	1	0	30.4
CC	0	0	6	3	0	1	2	23	4	0	0	7	1	1	1	1	2	1	3	4	4	0	0	0	4	1	0	4	7	1	1	2	13	4	0	1	0	0	0	29.4
EC	0	0	5	2	0	0	2	23	8	1	0	3	2	2	1	0	2	1	3	5	5	1	0	1	4	1	1	2	7	0	0	1	17	8	3	0	0	0	0	28.9
BC	0	0	6	3	0	1	3	21	1	0	0	5	0	1	1	0	2	0	3	4	4	0	0	0	2	1	1	3	5	1	1	2	14	5	0	1	0	0	0	28.2
Euclidean A	0	0	5	3	0	1	2	24	2	1	0	3	0	1	2	0	2	0	5	4	6	1	0	0	3	1	0	5	7	0	0	1	18	6	2	0	0	5	0	27.9
NC(RSL=0.99)	1	0	7	3	0	1	3	25	2	0	0	8	0	1	1	1	1	1	3	4	4	0	0	0	2	1	0	4	7	1	1	1	8	3	0	1	0	0	0	27.6
ED_5	1	0	7	3	0	1	3	25	2	0	0	8	0	1	1	1	1	1	3	4	4	0	0	0	2	1	0	5	6	1	1	1	8	3	0	1	0	0	0	27.5
$ED_{1/5}$	0	0	6	3	0	1	2	26	2	0	0	5	0	1	1	0	1	0	3	4	4	0	0	0	1	1	0	3	6	1	1	1	12	5	1	1	0	0	0	27.4
PR	0	0	5	5	0	1	3	23	3	0	1	2	2	2	1	0	2	0	3	4	4	0	1	0	3	1	0	3	7	1	1	1	17	5	0	0	0	0	0	27.3
NC(RSL=0.5)	0	0	5	3	0	1	3	24	2	0	0	7	0	1	1	1	1	1	3	4	4	0	0	0	2	1	0	3	7	1	1	1	12	4	0	1	0	0	0	26.5
ED_1	0	0	6	3	0	1	3	22	2	0	0	5	0	1	1	0	1	0	3	4	4	0	0	0	1	1	0	3	5	1	1	1	11	3	0	1	0	0	0	25.9
$ED_{1/2}$	0	0	5	3	0	1	1	21	1	1	0	4	0	1	1	0	0	0	3	3	4	2	0	0	1	1	0	0	6	1	1	0	14	5	2	1	0	0	0	25.2
Euclidean B	0	0	2	4	0	1	4	20	1	4	0	4	0	3	2	1	0	0	2	0	2	0	0	0	0	0	0	7	3	0	0	4	7	3	0	0	1	5	0	17.5

(b) Weighted edges by days

	A																								B	C	D			E		F				G	H	I	J	
	29	80	121	129	137	147	155	157	163	171	176	180	244	288	299	304	307	311	329	330	331	349	352	354	244	2	12	14	16	7	5	28	33	35	40	41	4	309	1	AST
ED_5	0	0	6	3	0	1	3	28	2	0	0	8	0	1	1	1	1	1	3	4	4	0	0	0	2	1	0	4	7	1	1	1	8	4	0	1	0	0	0	28.5
BC	0	0	6	3	0	1	3	18	4	0	0	4	1	1	1	0	2	0	3	4	4	0	0	0	2	1	0	4	5	1	1	2	12	6	1	1	0	0	0	26.3
NC(RSL=0.5)	0	0	8	2	0	1	6	16	1	0	1	1	0	4	2	0	2	0	5	2	6	0	0	0	2	1	3	3	7	0	0	0	12	5	2	0	1	6	1	26.3
$ED_{1/2}$	0	1	8	3	0	1	3	21	1	1	1	2	0	5	1	0	0	0	3	2	4	1	0	0	4	1	2	0	7	1	1	1	11	6	2	0	1	1	1	25.9
ED_1	0	0	6	4	0	2	2	27	2	0	1	4	0	3	2	0	0	0	5	3	6	0	0	0	1	1	2	2	6	1	1	0	6	4	0	0	1	6	0	24.9
PR	0	0	5	4	0	1	2	14	7	0	0	2	0	2	0	0	0	0	1	2	1	0	1	0	4	1	1	2	7	1	1	1	13	5	0	0	0	0	0	23.4
CC	0	0	5	4	0	1	2	17	11	0	1	3	1	3	2	0	2	0	3	3	4	0	0	0	2	1	0	2	6	1	1	0	6	6	0	3	1	0	0	22.0
NC(RSL=0.99)	0	1	3	2	0	2	5	16	0	0	1	0	0	4	0	0	2	0	2	0	2	0	0	0	3	0	3	1	4	1	1	2	13	6	2	0	1	6	1	19.4
DC	0	1	1	2	0	1	2	16	9	0	1	0	0	4	1	0	2	0	0	0	0	0	0	0	2	0	2	1	5	1	1	2	12	9	0	2	1	0	0	19.0
EC	0	1	3	1	0	1	2	10	0	0	1	0	0	3	1	0	2	0	0	0	0	0	0	0	2	0	2	1	2	1	1	1	8	8	0	2	1	1	1	17.7
Euclidean A	0	1	2	4	1	3	3	18	2	0	1	1	0	5	1	0	2	0	2	0	2	0	0	0	1	0	3	5	4	5	1	3	9	4	2	0	1	7	1	17.5
ED_2	0	0	4	3	1	4	4	15	1	2	1	3	0	4	2	1	0	0	2	0	2	0	0	0	1	0	3	3	5	4	0	4	6	2	2	0	1	7	1	17.0
$ED_{1/5}$	0	0	3	3	1	2	5	14	1	0	1	1	0	4	1	0	0	0	2	0	2	0	0	0	1	0	3	5	3	4	0	2	6	2	2	0	1	7	1	16.0
NC(RSL=0.1)	0	0	4	4	1	4	4	18	1	4	1	4	0	5	2	1	0	0	2	0	2	0	0	0	2	0	1	5	3	4	0	4	5	2	0	0	1	7	1	15.0
Euclidean B	0	0	2	4	1	4	1	11	2	4	0	3	0	3	2	1	0	0	2	0	2	0	0	0	0	0	0	7	3	4	0	4	6	3	0	0	0	6	0	12.6

(c) Weighted edges by hours

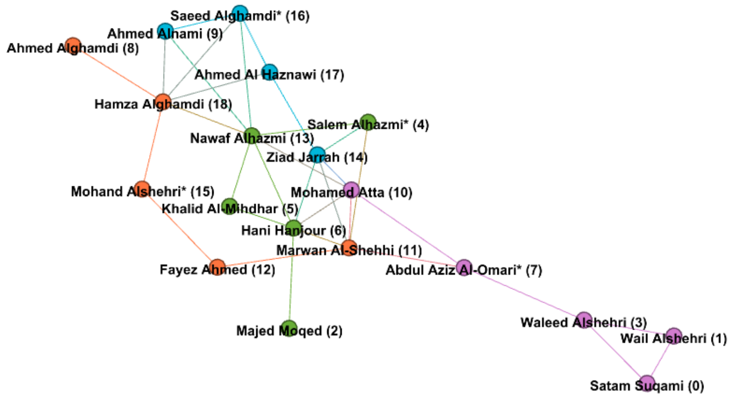
	A																								B	C	D			E		F				G	H	I	J	
	29	80	121	129	137	147	155	157	163	171	176	180	244	288	299	304	307	311	329	330	331	349	352	354	244	2	12	14	16	7	5	28	33	35	40	41	4	309	1	AST
ED_5	0	0	6	3	0	1	3	30	2	0	0	8	0	1	1	1	1	1	3	4	4	0	0	0	2	1	0	4	6	1	1	1	9	3	0	1	0	0	0	27.6
BC	0	0	5	3	0	1	2	28	2	0	0	2	0	2	0	0	0	0	1	1	2	0	1	0	5	1	2	2	7	1	1	1	13	6	1	0	0	0	0	26.6
PR	0	0	6	3	0	1	3	25	6	0	0	4	1	2	1	0	2	0	3	5	7	0	0	1	3	1	0	3	6	1	1	3	14	6	0	1	0	0	0	26.3
$ED_{1/2}$	0	1	8	4	0	1	3	19	1	1	1	3	0	5	1	0	0	0	3	3	4	1	0	0	3	1	2	1	6	1	1	1	12	6	2	0	1	1	1	25.7
ED_1	0	0	7	6	1	3	3	21	2	0	1	2	0	5	2	0	0	0	5	2	6	0	0	0	3	1	1	5	8	5	1	2	7	2	0	2	1	7	1	24.6
CC	0	1	7	4	0	5	7	15	2	0	1	3	1	4	0	0	2	0	2	0	3	0	0	0	3	1	2	1	4	9	2	1	9	5	2	1	1	1	1	21.4
DC	0	1	3	2	0	1	2	15	0	0	1	0	0	5	0	0	0	0	0	0	0	0	0	0	2	0	3	1	2	1	1	4	15	3	2	0	1	1	1	18.7
NC(RSL=0.99)	0	0	3	4	1	2	4	14	1	0	1	1	0	5	1	0	0	0	2	0	2	0	0	0	2	0	3	5	3	4	0	3	8	2	2	0	1	7	1	18.2
NC(RSL=0.5)	0	0	3	3	1	4	4	20	1	0	1	2	0	4	1	0	0	0	2	0	2	0	0	0	1	0	2	5	3	4	0	2	6	1	2	0	1	7	1	17.9
Euclidean A	0	0	3	3	1	4	1	16	1	0	1	2	0	5	0	0	0	0	2	0	2	0	0	0	1	0	2	5	4	4	0	3	11	2	2	0	2	7	1	17.5
EC	0	1	3	2	0	1	2	8	0	0	1	0	0	5	0	0	0	0	0	0	0	0	0	0	2	1	3	0	2	1	1	3	14	5	3	0	1	1	1	17.3
ED_2	0	0	2	4	1	4	1	16	1	4	1	4	0	4	2	1	0	0	2	0	2	1	0	0	1	0	1	5	2	4	0	4	7	1	2	0	1	7	1	16.6
$ED_{1/5}$	0	0	3	3	1	4	1	15	1	2	1	3	0	3	2	1	0	0	2	0	2	1	0	0	1	0	1	5	2	4	0	4	7	1	2	0	1	6	2	16.4
Euclidean B	0	0	1	5	1	4	1	18	1	4	0	3	0	3	2	1	0	0	2	0	2	0	0	0	1	0	0	5	2	5	1	2	9	3	0	0	0	7	0	15.8
NC(RSL=0.1)	0	0	3	3	2	2	5	16	3	5	1	4	0	4	2	1	0	0	2	0	2	1	0	0	2	0	0	3	2	6	0	2	5	3	0	0	1	7	1	15.5

Source: Andrade et al. [73].

4.4.3 Network of hijackers in the September 11 attacks

The network of hijackers responsible for the 9/11 operation is the same studied in [28]'s work and is displayed in Figure 22. This network was built with network data collected by [86], the nodes represent the hijackers and the links were established given by living together or attending the same school, or the same classes/training, or traveling together and attending meetings together. The network has a single component formed by 19 nodes and 33 edges, has a density equal to 0.193 and the average clustering coefficient is 0.475. The colors in the network refer to the different flights of American Airlines (AA) and United Airlines (UA) that were used by the terrorists; i.e., AA-77 (green),AA-11 (lilac), UA-93 (blue) and UA-175 (orange).

Figure 22 – Operational network of hijackers of Al Qaeda's 9/11 attack. AA-77 (green), AA-11 (lilac), UA-93 (blue) and UA-175 (orange).

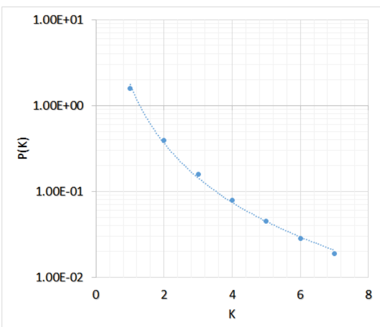


Source: Lindelauf et al. [28].

The nodes' weights were defined by Lindelauf et al. [28] based on reports and historical sources, followed some indicators: attending meetings on terror attack planning, affiliation, accomplice to previous attacks, attending terrorist training camps, and signs of radicalization. The authors first assigned weight 1 to all of the 19 nodes in the network and then for each indicator added 1 to a hijacker's weight if that hijacker participated in the activity related to that indicator. Thus, Nodes 10 and 14 obtained a total weight of 4; nodes 5 and 11, a total weight of 3, node 13, a total weight 2; and the total weight of the remaining nodes were equal to 1.

The degree distribution of the network shown in Figure 23 present power law distributions.

Figure 23 – Degree distributions (p-value = 1.01E-24), where the p-value was from the Kolmogorov-Smirnov test. Note that the vertical axis is on a logarithmic scale.



Source: Andrade et al. [73].

4.4.3.1 Attribute load

The nodes' attributes in the hijackers network were obtained through indicators, individuals with high scores are considered fundamental for the success of the operation.

Featuring a single component, the attribute load brings together the importance of the 19 individuals involved in the operation.

Figure 24 shows how each targeting method under analysis causes the loss of the attribute after the removal of q_{max} nodes. The measures are listed in increasing order of $A(n)$, which is the average relative loss of attribute load when all nodes are removed according to the given targeting method and is displayed in the right side of the figure. The smaller this average, the greater the efficiency of the measure in removing first the most important nodes in the network.

Note that the most efficient measures are those calculated including the attribute of the nodes. ED_1 is the most efficient and ED_2, ED_5, $ED_{1/2}$ and NC(RSL=0.5) generated the second best result. Traditional measures of centrality were the least efficient in reducing the attributes load on the network of hijackers.

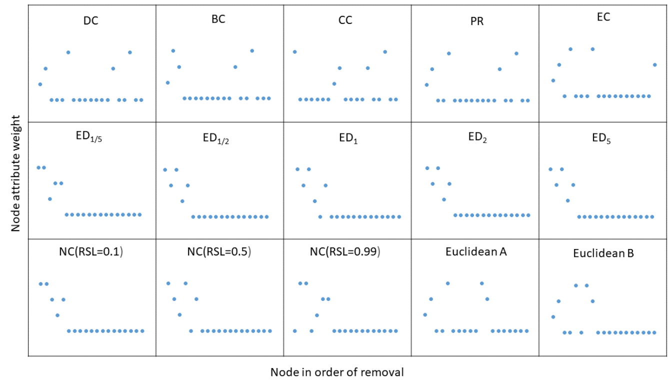
Figure 24 – Attribute load as a function of the total number of nodes removed q_{max}.

k_q/k \ q_{max}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	A(n)
ED_1	0.87	0.63	0.50	0.23	0.17	0.13	0.10	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.17222
ED_2	0.87	0.63	0.50	0.23	0.23	0.13	0.10	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.17593
ED_5	0.87	0.63	0.50	0.23	0.23	0.13	0.10	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.17593
$ED_{1/2}$	0.87	0.63	0.57	0.23	0.17	0.13	0.10	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.17593
NC(RSL=0.5)	0.87	0.63	0.50	0.23	0.23	0.13	0.10	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.17593
NC(RSL=0.1)	0.87	0.73	0.50	0.23	0.23	0.13	0.10	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.18148
$ED_{1/5}$	0.87	0.73	0.67	0.23	0.23	0.13	0.10	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.19074
NC (RSL=0.99)	0.87	0.73	0.60	0.53	0.37	0.17	0.10	0.10	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.21667
Euclidean B	0.93	0.83	0.70	0.53	0.27	0.17	0.10	0.10	0.07	0.07	0.07	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.23148
BC	0.93	0.80	0.43	0.30	0.30	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.07	0.03	0.03	0.03	0.23519
Euclidean A	0.93	0.83	0.70	0.53	0.27	0.17	0.17	0.13	0.13	0.13	0.10	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.24074
EC	0.93	0.83	0.70	0.43	0.33	0.17	0.17	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.25926
PR	0.93	0.83	0.70	0.53	0.27	0.27	0.17	0.17	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.03	0.03	0.03	0.27778
DC	0.93	0.83	0.70	0.53	0.43	0.17	0.17	0.17	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.03	0.03	0.03	0.28148
CC	0.87	0.80	0.77	0.73	0.37	0.37	0.27	0.27	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.03	0.03	0.03	0.30926

Source: Andrade et al. [73].

It can be seen from Figure 25 that there is no relationship between the attribute and the centrality of the node. Therefore, removing the most central nodes does not necessarily remove the nodes with the greatest contribution to the success of the operation. On the other hand, the most efficient measures in reducing the load of attributes were those that are able to remove nodes with higher attribute weights first.

Figure 25 – Scatterplot of node removed and node attribute weighting.



Source: Andrade et al. [73].

4.4.3.2 Robustness

In the robustness analysis, Figure 26, ED_1 and NC(RSL=0.5) are the most efficient measures in reducing the size of the component, followed by ED_2, ED_5 and $ED_{1/2}$ that showed the same efficiency. Just like in the anklets network, the best attack strategy was not entirely a topological intervention, but a measure that mixes characteristics of the network topology with nodes' attributes.

Figure 26 – Robustness as a function of the total number of nodes removed q_{max}.

n_q/n \ q_{max}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	R(n)
ED_1	0.95	0.68	0.63	0.37	0.21	0.21	0.16	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.21930
NC(RSL=0.5)	0.95	0.68	0.63	0.37	0.21	0.21	0.16	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.21930
ED_2	0.95	0.68	0.63	0.37	0.37	0.21	0.16	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.22807
ED_5	0.95	0.68	0.63	0.37	0.37	0.21	0.16	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.22807
$ED_{1/2}$	0.95	0.68	0.63	0.37	0.37	0.21	0.16	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.22807
Euclidean A	0.95	0.89	0.68	0.53	0.26	0.16	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.23684
NC(RSL=0.1)	0.95	0.89	0.63	0.37	0.37	0.21	0.16	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.23977
NC (RSL=0.99)	0.79	0.74	0.68	0.58	0.37	0.16	0.16	0.16	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.23977
$ED_{1/5}$	0.95	0.89	0.84	0.37	0.37	0.21	0.16	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.25146
PR	0.95	0.84	0.68	0.53	0.26	0.21	0.16	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.25439
Euclidean B	0.95	0.89	0.68	0.53	0.26	0.16	0.16	0.16	0.11	0.11	0.11	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.25731
DC	0.95	0.79	0.74	0.53	0.26	0.21	0.21	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.26023
CC	0.95	0.89	0.63	0.47	0.37	0.21	0.21	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.11	0.11	0.05	0.05	0.29240
BC	0.95	0.74	0.68	0.63	0.63	0.42	0.26	0.16	0.16	0.16	0.11	0.11	0.11	0.11	0.05	0.05	0.05	0.05	0.05	0.30117
EC	0.95	0.79	0.74	0.47	0.37	0.21	0.21	0.21	0.21	0.21	0.21	0.16	0.16	0.16	0.16	0.16	0.11	0.05	0.05	0.30702

Source: Andrade et al. [73].

4.4.3.3 Toughness

In the analysis of toughness, it is realized that although ED_1, $ED2$ and ED_5 remove nodes faster from the largest component, the number of edges lost after removing the nodes is less than the number of edges lost after removing nodes by the less efficient measures in terms of robustness: EC, BC and DC. NC(RSL=0.5) presented good results in the three analyzes, attribute load, robustness and toughness.

Figure 27 – Toughness as a function of the total number of nodes removed q_{max}.

W_q/W	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	$R_w(n)$
EC	0.79	0.61	0.42	0.24	0.12	0.06	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12963
Euclidean A	0.79	0.61	0.42	0.30	0.15	0.06	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13468
NC(RSL=0.5)	0.85	0.58	0.39	0.24	0.12	0.12	0.06	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13636
BC	0.79	0.64	0.24	0.24	0.09	0.09	0.09	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.00	0.00	0.00	0.13636
DC	0.79	0.61	0.42	0.30	0.18	0.06	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13636
PR	0.79	0.61	0.42	0.30	0.15	0.15	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13973
ED_1	0.85	0.58	0.48	0.24	0.12	0.12	0.06	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14141
Euclidean B	0.79	0.61	0.42	0.30	0.15	0.06	0.06	0.06	0.03	0.03	0.03	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.14646
ED_2	0.85	0.58	0.48	0.24	0.24	0.12	0.06	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14815
ED_5	0.85	0.58	0.48	0.24	0.24	0.12	0.06	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14815
$ED_{1/2}$	0.85	0.58	0.48	0.24	0.24	0.12	0.06	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14815
NC(RSL=0.1)	0.85	0.73	0.48	0.24	0.24	0.12	0.06	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15657
$ED_{1/5}$	0.85	0.73	0.55	0.24	0.24	0.12	0.06	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15993
NC (RSL=0.99)	0.82	0.67	0.58	0.39	0.18	0.09	0.09	0.09	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16667
CC	0.85	0.70	0.61	0.52	0.27	0.18	0.12	0.06	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18687

q_{max}

Source: Andrade et al. [73].

4.4.3.4 Analysis of the order of removal of nodes.

Analyzing the different rankings in Table 2 leads to the following observations. ED_1 and NC(RSL=0.5), the first and second most efficient measure in terms of attribute loss and reduction of network robustness, are able to remove hijackers from each of the four different forms of flight in the first 4 removals. The fifth removal following these measures as attack strategies eliminates all communication between hijackers from different flights.

Note also that Nawaf Alhazmi (13) is the first individual chosen by most centrality measures and Mohamed Atta (10) is the first chosen one by ED measures. Mohamed Atta has a greater contribution to the network, as it has a greater attribute weight. Mohamed Atta was designated as the ringleader of the 9/11 hijackers according to members of the 9/11 Commission Report, [105], and by Osama bin Laden's statements. Neither of them causes a disruption of the network, although Mohamed Atta is more important, as it intermediates the AA-11 hijackers with the others, being a fundamental communication link. On the other hand, Nawaf Alhazmi has more toughness, as it has a greater number of connections. Marwan Al-Shehhi (11) is the second most rated for the second removal. He and Mohamed Atta dismember AA-11 from the network.Based on ED measures, the first four nodes removed comprise all four pilots from the four different flights.

Table 2 – Rankings for Al Qaeda's 9/11 network based on the measurements used.

DC	13	11	18	6	3	14	16	0	7	12	1	2	4	5	8	9	10	15	17
BC	13	14	11	7	18	6	16	0	1	2	3	4	5	8	9	10	12	15	17
CC	10	6	4	17	12	18	3	13	11	0	9	1	2	5	7	8	14	15	16
PR	13	11	18	6	14	16	3	0	7	12	1	2	4	5	8	9	10	15	17
EC	13	11	18	14	3	6	16	10	0	15	12	17	4	2	9	8	7	1	5
$ED_{1/5}$	10	14	13	11	5	18	3	16	15	6	1	17	12	9	8	7	4	2	0
$ED_{1/2}$	10	11	14	13	5	18	3	16	15	6	1	17	12	9	8	7	4	2	0
ED_1	10	11	14	13	18	5	3	16	15	6	1	17	12	9	8	7	4	2	0
ED_2	10	11	14	13	5	18	3	16	15	6	1	17	12	9	8	7	4	2	0
ED_5	10	11	14	13	5	18	3	16	15	6	1	17	12	9	8	7	4	2	0
NC(RSL=0.1)	10	14	11	13	5	18	3	16	15	6	1	17	12	9	8	7	4	2	0
NC(RSL=0.5)	10	11	13	14	18	5	3	16	15	6	1	17	12	9	8	7	4	2	0
NC(RSL=0.99)	7	14	10	18	13	11	5	16	3	15	6	1	17	12	9	8	4	2	0
Euclidean_A	13	11	18	6	14	3	16	7	0	12	10	5	1	2	4	8	9	15	17
Euclidean_B	13	11	18	6	14	3	10	5	16	1	2	4	8	9	15	17	7	0	12

Source: Andrade et al. [73]

4.5 CONCLUSIONS

The application in an anklet network, whose connections are formed by encounters and the attributes considered were extracted from displacement data, shows that the targeted attack following the proposed measure, when parameter t is equal to 2 or 5, is more efficient than the other measures. Throughout the process of removing the nodes, the attribute load, the robustness and the toughness of the network, targeting according to the ED_5 and ED_2, remained lower than the damage caused by other targeted measures. In the Al Qaeda's 9/11 network, although ED_2 and ED_5 had good results, ED_1 was more efficient. However, ED measures in this network were not as efficient in reducing toughness. It is view as a surprise that ED, which is not a purely topological measure, is more effective in identifying nodes whose removal affects the network structure than centrality measures, which are exclusively topological in nature.

The measure proposed was also more efficient than the NC measure and the Euclidean distance, both which also includes the nodes' attributes and they have been proposed as targeting measures to dismantle or disrupt criminal network both structurally due the connections and interactions of the actors and functionally, limiting the resources or attributes that actors have. It is important to report that for values of t greater than 5 the degradation results in the network are similar to those presented by t equal to 5. As the distance between nodes approaches 1 when t grows, it can be said that the distance between nodes becomes not be a determining factor in the choice of degrading nodes. Moreover, the proposed measure tends to capture the most cohesive nodes who occupy strategic brokerage positions. A broker is a node that acts as a bridge between two groups of nodes that are disconnected with its removal [106]. Members of these two groups depend on the broker for indirect access to resources beyond their reach [98].

In the analysis of the crimes, it was found that Art. 157 of the penal code and Art. 33 of the narcotics law are the most violated by the first 20 nodes targeted by the attack. In this same analysis, it was noted that the measures derived from ED_t were able to remove those convicted with a longer average sentence, ED_2 in the unweighted network and ED_5 in the weighted network. It was noted that the 20 nodes removed, considering all targeting methods, from the unweighted network have an average longer condemnation time, which is about 7 years longer than in the day-weighted network and more than 7 years than in the hour-weighted network. This small sample indicates that the strength of the links does not corroborate for the capture of nodes with a longer average condemnation time. There are also indications that the greater heterogeneity of the strengths of the links and the greater amplitude (based on the Figure 16) makes the network weighted by the hour the least efficient in removing, in the first attacks, the nodes with longer average condemnation times. In the Al Qaeda's 9/11 network, the ED measures first eliminate the pilots of the four flights, these hijackers were instrumental in the operationalization of the attack. It is noteworthy that the first pilot removed, Mohamed Atta, was considered the leader of the hijackers.

Future research may analyze the effectiveness of police action, imprisonment, in removing convicts from the network in order to avoid a possible criminal act. Here it is tested the strategy of removing 20 convicts with greater recurrence in the crime of armed robbery. The results show that this is not an efficient strategy. The removal of a single node according to the most efficient measure of targeted attack, ED_5, ED_2 and BC, caused greater damage to the network. On the other hand, if an efficient centrality measure is not used as a targeting method, its damage to the network can be lower than simply removing the 20 convicts with greater recurrence in the crime of armed robbery. In the applications, this is the case for $NC(RSL = 0.1)$, $ED_{1/2}$ and $ED_{1/5}$.

5 A PROPOSAL FOR THE EI INDEX FOR OVERLAPPING GROUPS

Some results discussed in this chapter was published on The 8th International Conference on Complex Networks and Their Applications (Andrade and Rêgo [107]).

5.1 CONTEXT

Individuals have similar attributes and similar relationships that can lead to the formation of communities [108]. Several studies, based on real-world's data, are interested in networks in which the nodes are assigned to groups whose similarities have been traditionally related to with aspects of personal features or personal attributes [109]. For example: age-based [54]; based on ethnicity [55]; based on gender [110], religion, politics [111]; among others. Grouping involves partitioning the set of nodes into exhaustive and mutually exclusive subsets. Publications that use the EI index as a measure of homophily or segregation are concentrated in disjoint or mutually exclusive groups, that is, each node or actor in the network has only one bond of a particular attribute (for example, the sex attribute, where a node has either a masculine or feminine sex) and thus is inserted in only one group (for example group of nodes whose sex is masculine or feminine). By the characteristics of the attributes, it is possible to generate only disjoint groups: sex (masculine, feminine), age (greater than "x" or less than "x", for example), religion (Catholic, Spiritist, Protestant), like political option, social level, degree of instruction, among other attributes.

Therefore, situations where network actors are present in more than one group, such as non-disjoint groups, are not commonly explored. However, Lee and Brusilovsky [109] points out that society is currently goaded by information and knowledge and conducts to new dimensions of homophily. Information, knowledge and some attributes such as: economic blocks in commercial networks; communities in social networks such as Facebook, Twitter, among others; and other attributes linked to behaviors, tastes and attitudes generate non-disjointed groups. One of the barriers encountered in the analysis of non-disjointed groups is the absence of a measure, since the EI index is defined for disjoint groups.

In this context, the objective of this work is to develop a new measure that quantifies the relational structure within and between groups that encompass not only the analysis of disjoint groups but also non-disjoint groups. Allowing the expansion of the analysis of social networks, for several types of attributes, and thus generate previously unexploited knowledge. Specifically, it is generalized the EI index developed by Krackhardt and Stern [57]. As far as it is aware of, the unique work that mentions about EI index for non-disjoint groups is that of [34]. However, in this work, the authors do not formalize this definition and do not apply the metric in any network. They even state that the EI index for the overall network would pose some kind of problem. So, how can the influence of similarity or dissimilarity be analyzed in the relationships between individuals who have attributes related to more than one group? This work fills the literature gap by providing an EI index definition for arbitrary sets of nodes (which include unitary sets and also the whole set of nodes) and for group attributes.

This chapter is organized as follows. In Section 5.2, the measure proposed is presented, which is a generalization of the current EI index, encompassing non-disjoint groups. Two applications of the proposed measure are made in Section 5.3. Finally, it is discussed the results of the applications in Section 5.4 and present conclusions.

5.2 EI INDEX: OVERLAPPING CASE

Krackhardt and Stern designed the EI index of friendships between members of an organization, both internal and external, to organizational sub-units on crisis environment. Sub-units are disjoint groups, so the measure was initially designed to analyze disjoint groups. Presently, publications that use the EI index as a measure of homophily or segregation are concentrated into disjoint or mutually exclusive groups, i.e., each node or actor in the network has only one attribute in common and thus is inserted in only one group. However, distinctive disjoint groups rarely exist at large scales in many empirical networks [36].

In a literature search, the unique work that mentions about calculating EI index for non-disjoint groups is Everett and Borgatti [34]. However, this is not the focus of their chapter and they do not provide a clear explanation about what they had in mind. In fact, they state that: “...in calculating a score for any individual node, the EI measure classifies all ties as in-group ties or out-group ties, effectively forming a temporary partition." From this statement, it is not clear how they propose to classify a tie as in-group or out-group. For example, one can classify a tie as in-group if both nodes belong to at least one same group or if both nodes belong to exactly the same groups. For the group level, the authors complete: “If a group-level index is needed, the internal and external counts can be aggregated for that group, and an overall EI score computed for the group”. Once again, they are not clear about how to aggregate internal and external counts for a group. In this sense, in order to explore cases of non-disjoint groups, it is developed a new method to obtain the EI index, which is a generalization of the current method which is designed for disjoint groups.

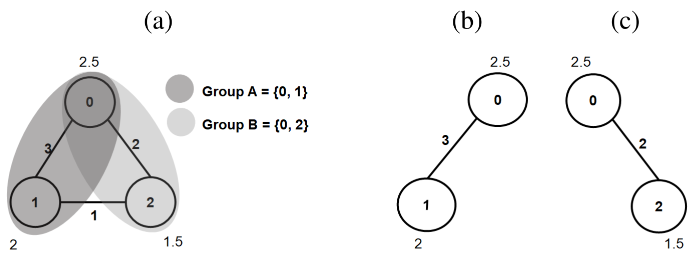
Before giving the formal definitions, first it is used the graphs displayed in Figure 28 to illustrate the proposed definitions. The graph in Figure 28a has three nodes and two generic attribute groups, nodes 0 and 1 have attribute A and nodes 0 and 2 have attribute B. Therefore, both attribute groups have one node in common. In the graphs shown in Figures 28b and 28c, nodes are connected if they are connected in the graph shown in Figure 28a and belong to the same group.

In this context, the EI index is defined as follows:

(i) For a set of nodes: EL is the number of edges linking at least one node in the given set that are in the graph shown in Figure 28a but are not present in any of the group graphs shown in Figures 28b and 28c and IL is the number edges connecting at least one node in the given set that are in the graph shown in Figure 28a and are also present in at least one of the group graphs shown in Figures 28b and 28c.

(ii) For an attribute group: EL is the number of edges linking at least one node with a given attribute group (A or B) that are in the graph shown in Figure 28a but are not present in the given group graph shown in Figure 28b or 28c and IL is the number of edges connecting at least one node with a given attribute group (A or B) that are in the graph shown in Figure 28a and are also present in the given group graph shown in Figure 28b or 28c.

Figure 28 – Non-disjoint group example



Source: The Author (2021)

Case (i) resembles the traditional EI index measure. First, any individual edge is classified as internal, if the corresponding pairs of nodes have some attribute group in common, or external, otherwise. Then, for any set of nodes, the EI index is obtained by summing the counts of internal and external edges linking at least one node in the set. This case includes both sets of single nodes and the entire network which is seen as the set of all nodes. Case (ii) is used specifically to find the EI index of a given attribute group. Here an edge is internal if it connects two nodes having the given group attribute and external, otherwise.

Let us now formally define the EI index for non-disjoint groups. Let A be the set of all attributes for nodes in a social network with n nodes. For $X \in A$, let $I_X(v_i)$ be equal to 1 if node v_i has attribute X and 0, otherwise. Moreover, for a generic set of nodes, S, consider the following sets of indices $I(S) = \{i : v_i \in S\}$ and $J^i(S) = \{j : (v_j \in S \text{ and } j > i) \text{ or } (v_j \notin S \text{ and } j \neq i)\}$. Thus, the number of external and internal links is given, respectively, by:

$$EL(S) = \sum_{i \in I(S)} \sum_{j \in J^i(S)} x_{ij}(1 - \max_{X \in A}\{I_X(v_i)I_X(v_j)\})$$

and

$$IL(S) = \sum_{i \in I(S)} \sum_{j \in J^i(S)} x_{ij} \max_{X \in A}\{I_X(v_i)I_X(v_j)\},$$

where in the unweighted case x_{ij} is either 1 or 0 depending on whether there is a link between node v_i and v_j, in the case with only edges' weights $x_{ij} = w_{ij}$ and in the case of edges' and nodes' weights $x_{ij} = z_{ij}$.

Alternatively, for $X \in A$, it can be defined the number of external and internal links for the group of nodes, S_X, that has attribute X, respectively, as follows:

$$EL(S_X) = \sum_{i=1}^{n} \sum_{j>i} x_{ij} I_X(v_i)(1 - I_X(v_j))$$

and

$$IL(S_X) = \sum_{i=1}^{n} \sum_{j>i} x_{ij} I_X(v_i) I_X(v_j),$$

where x_{ij} is defined exactly as before.

Table 3 displays the results for the graph shown in Figure 28. It is easy to verify that the proposed metric is a generalization of the EI index proposed in [57] in the sense that if groups are disjoint, then it coincides with (Equation 17).

It is important to highlight that unlike the case for disjoint groups, in the case of non-disjoint groups the EI index for a given group attribute can be different from the EI index for the set of nodes which have this given attribute. For example, in Table 3, one can see that the EI index for Group A is different from the EI index of the set $\{0, 1\}$ and that of Group B is different from that of the set $\{1, 2\}$.

Table 3 – EI Index non-disjoint group example

	Unweighted			Weighted		
	EL	IL	EI Index	EL	IL	EI Index
set {0}	0	2	-1.00	0	5	-1.00
set {1}	1	1	0.00	1	3	-0.50
set {2}	1	1	0.00	1	2	-0.33
set {0,1}	1	2	-0.33	1	5	-0.67
set {0,2}	1	2	-0.33	1	5	-0.67
set {1,2}	1	2	-0.33	1	5	-0.67
set {0,1,2}	1	2	-0.33	1	5	-0.67
Group A	2	1	0.33	3	3	0.00
Group B	2	1	0.33	4	2	0.33
	Z_unweighted			Z_weighted		
	EL	IL	EI Index	EL	IL	EI Index
set {0}	0	4.25	-1.00	0	10.75	-1.00
set {1}	1.75	2.25	-0.13	1.75	6.75	-0.59
set {2}	1.75	2	-0.07	1.75	4	-0.39
set {0,1}	1.75	4.25	-0.42	1.75	10.75	-0.72
set {0,2}	1.75	4.25	-0.42	1.75	10.75	-0.72
set {1,2}	1.75	4.25	-0.42	1.75	10.75	-0.72
set {0,1,2}	1.75	4.25	-0.42	1.75	10.75	-0.72
Group A	3.75	2.25	0.25	5.75	6.75	-0.08
Group B	4	2	0.33	8.5	4	0.36

Source: The Author (2021)

5.3 APPLICATION

In this section, the method proposed will be applied in two networks object of studies in previous publications. These networks present the fundamental element for approach this work, which is the presence of non-disjoint groups, in addition to information about the nodes' weights. In this way, the EI Index will be obtained for 4 situations: without considering the weight of the edges and the nodes, unweighted (UW); regarding only the nodes' weight, Z_unweighted (ZU); considering only the edges' weight, weighted (W); taking into account both weights, Z_weighted (ZW). To evaluate whether the EI Index for a given group is compatible with the expected when connections occur randomly, i.e., without preference of members for external or internal relations, for the unweighted and the Z_unweighted cases, it is calculated the expected EI index for each one of the groups analyzed considering the average of 5000 randomly generated binomial graphs with the same density as that of the original graphs. It is also added a probability, p-value, that expresses how unlikely it is to obtain an EI index at least as extreme as the one observed in the randomly generated binomial graphs. It is considered one sided p-values so that it is given by the relative frequency of times that the simulated EI has a value greater (resp., less) than or equal to the observed EI, when the expected EI is lees (resp., greater) than the observed one.

5.3.1 Data

In order to implement the proposed EI index, it is used data from two real networks. Next some details about these networks, it will be given.

(i) Co-authorship PQ: The PQ network is co-authorship network among researchers in the area of Production Engineering in Brazil that held a Productivity Research scholarship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) in 2015. It has 92 nodes in the giant component and 131 edges. The network is undirected and the edges' weights represent the number of publications made in co-authorship by a given pair of researchers in the period of 2005 to 2014 [71].

(ii) Trade of American Countries: The network of commerce between American countries is formed 30 countries and 356 edges. This network was created from the network of international trade developed by Andrade and Rêgo [14] which includes 178 countries from all continents, forming a unique main component with 10,419 edges. The network is undirected and the edges' weights represent the mean of export and import commercial transactions between a pair of countries in the period of 2015.

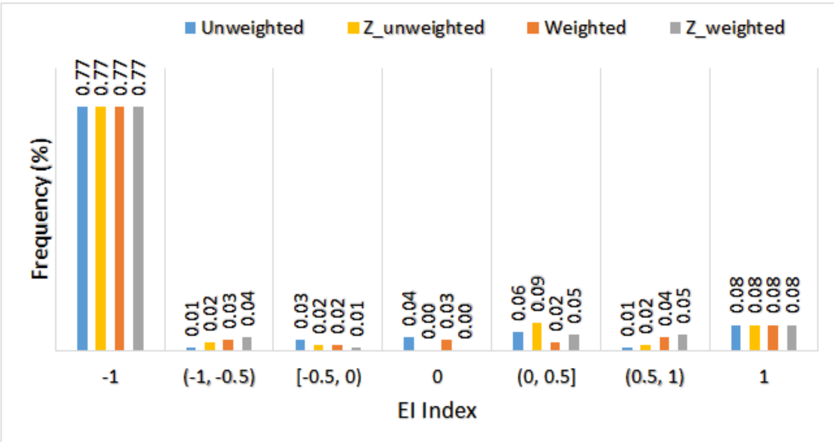
5.3.2 PQ Network

In order to calculate an EI index for each group and for each individual co-author in the network, it is necessary to partition nodes into groups. For the purposes of this study, it is divided the nodes into 9 groups that represent production engineering knowledge areas. These areas of knowledge are: Operations Engineering (OE), Operational Research (OR), Organizational Engineering (OrE), Economic Engineering (EE), Logistics (LO), Labor Engineering (LE), Product Engineering (PE), Sustainability Engineering (SE) and Quality Engineering (QE). It is uses the researchers' h-index as the nodes' weights. The h-index is a measure that combines, in a simple way, the quantity of publications and the impact of publications and is given by the maximum value of h such that a researcher has published h works and each one of these works have been cited h or more times [112].

Figure 29 shows how the relationships between researchers occur. It is clear that most researchers relate only to researchers who work in some common area, i.e., there is a prevalence of exclusively internal relationships. This result does not change when considering the frequency of the relationships or when considering the nodes' weights. Only 15% of researchers have both internal and external relationships and only 8% of scholars collaborate exclusively with scholars from areas other than their own.

For the whole network, the unweighted EI index is -0.74 ($p = 0.000$), the Z_unweighted EI index is -0.77 ($p = 0.000$), the weighted EI index is -0.74 and Z_weighted EI index -0.76. It comes as no surprise that, like the nodes, the whole network followed the same strong homophily trend.

Figure 29 – Distribution of EI index values in the PQ Network



Source: The Author (2021)

Table 4 presents the EI index of the researchers' knowledge areas, considering that researchers can act in more than one area. It can be noted that only researchers in the field of operational research cooperate more with each other than with scholars from other areas. This suggests that OR is a strong and independent research area. It is observed an even lower EI index when considering the nodes' and edges' weights, indicating a strong relationship between the members of the group and that OR researchers have a higher h-index than their external neighbors.

Logistics researchers have more external than internal links, given that the unweighted EI index is positive. It can be deduced that its external co-authors have a higher h-index, due to the increase of heterophily. On the other hand, when analyzing the frequency of connections, there is an inversion of preferences, logistics researchers cooperate more frequently with their internal neighbors.

Operations engineering has a balance between external and internal relations, however with a tendency to relate more frequently with external co-authors who have a high h-index.

Scholars in the field of economic engineering and also in sustainability engineering do not cooperate with others in their field. Both have totally heterogeneous relationships. These two areas of research, therefore, are strongly dependent on other areas.

Except for the cases of economic, product and sustainability engineering, all other areas of Production Engineering have a statistically significant tendency of being more connected to other researchers in the same field than in a random network. This result remains true both if researchers h-index are considered or not in the analysis.

Previous studies show that cooperation is a way to overcome the growing complexity and specialization of scientific research. By cooperating, each actor involved influences ideas, shares different experiences and acquires new skills, enabling the emergence of innovations more easily [113].

It is also analyzed the behavior of groups of researchers who exhibit the same scholarship level. The scholarship level in order of importance and the total number of researchers are: 1A (8%), 1B (5%), 1C (8%), 1D (19%) and 2 (59%). From Table 5, it is evident, and statistically significant, that for all levels of scholarships, researchers mostly collaborate with researchers who work in common areas. Level 1A and 1D researchers are the ones that most strengthen their research areas, while level 2 and 1C researchers are more diversified in collaboration areas. These relational patterns are not changed by including the edges' or nodes' weights. On the contrary, there is an intensification of internal relations, except for the 1D level.

Table 4 – EI indexes for Production Engineering areas in the PQ Network

Areas	UW	Expected	p-value	ZU	Expected	p-value	W	ZW
OR	-0.59	0.00	0.000	-0.64	-0.05	0.000	-0.62	-0.65
OE	0.05	0.65	0.000	0.11	0.63	0.000	0.15	0.17
LO	0.31	0.86	0.000	0.40	0.88	0.000	-0.30	-0.22
LE	0.54	0.94	0.000	0.39	0.94	0.000	0.43	0.32
OrE	0.56	0.91	0.000	0.58	0.92	0.002	0.70	0.66
QE	0.59	0.89	0.001	0.55	0.88	0.002	0.49	0.36
PE	0.83	0.93	0.100	0.82	0.93	0.108	0.91	0.88
EE	1.00	0.95	0.645	1.00	0.97	0.642	1.00	1.00
SE	1.00	0.97	0.855	1.00	0.98	0.852	1.00	1.00

Source: The Author (2021)

Table 5 – EI indexes for scholarship level groups considering the Production Engineering areas as class attributes in the PQ Network

Scholarship	UW	Expected	p-value	ZU	Expected	p-value	W	ZW
Level 2	-0.66	-0.05	0.000	-0.71	-0.10	0.000	-0.69	-0.70
Level 1D	-0.87	-0.18	0.000	-0.77	-0.20	0.000	-0.73	-0.76
Level 1C	-0.64	-0.22	0.000	-0.65	-0.21	0.000	-0.65	-0.65
Level 1B	-0.73	-0.16	0.000	-0.77	-0.19	0.000	-0.72	-0.77
Level 1A	-0.87	-0.30	0.000	-0.91	-0.36	0.000	-0.91	-0.91

Source: The Author (2021)

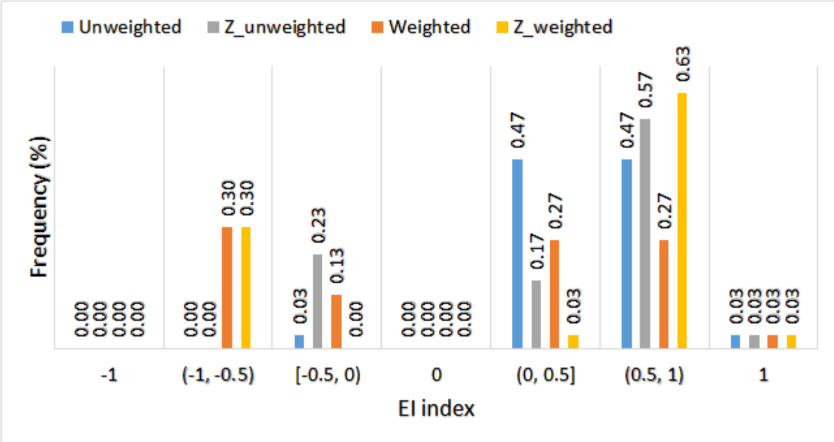
5.3.3 Trade of American Countries Network

In order to calculate an EI index for each attribute group and for each individual co-author in the network, it is necessary to first define the attribute groups. For the purposes of this study, it was defined 8 groups that represent the trade agreements among countries of the American continent. The trade agreements are: Latin American Integration Association (LAIA), Andean Community (AC), Caribbean Community (CARICOM), Central American Integration System (CAIS), Southern Common Market (SCM), North American Free Trade Agreement (NAFTA), Organisation of Eastern Caribbean States (OECS) and Central America Free Trade Agreement and Dominican Republic (CAFTA-RD). The gross domestic product (GDP) was defined as the node's weights.

In Figure 30, it can be seen that none of the countries of the American continent establishes relations exclusively with the economic block to which it belongs. Furthermore, trade relations external to the block are superior to internal relations, for more than 90% of the countries. By including the strength of relationships: Z_unweighted, weighted and Z_weighted case, it was observed a fluctuation in heterophily relationships and a small increase in homophily relationships.

The whole trade network is much more externally oriented, the unweighted EI index is 0.43 ($p = 0.009$). The network maintains the trend when considering the nodes' weight, but the observed Z_unweighted EI index of 0.43 is not statistically significant ($p = 0.504$), implying that in this case it cannot be disregarded that links are randomly made with respect to existence of trade agreements. An inversion of preferences is observed when analyzing the transactional volume, now the entire network is more suitable toward in-blocks ties, weighted EI index is -0.69 and Z_weighted EI index -0.72. This same pattern is observed in the distribution of EI index values in the network, as shown in Figure 30.

Figure 30 – Distribution of EI index values in Trade Network



Source: The Author (2021)

Analyzing the relations among economic blocks displayed in Table 6, it is seen that all of them present external relations superior to internal ones and that the OECS block has only external relations. However, it cannot be concluded that the trade agreements have not been put in place, since the observed EI index are not significantly statistically different from the EI indexes obtained in a random network, except for the cases of the Caribbean community and the Organisation of Eastern Caribbean States agreements, where countries have a statistically higher chance of trading with external countries than in a random network.

When including the nodes' weights, it was observed that the external relations superior to the internal ones are still within the expected. Nevertheless, by including the volume traded, only the NAFTA economic block shows a tendency toward in-blocks ties.

Table 6 – EI indexes for Trade agreement blocks in the Trade Network

Blocks	UW	Expected	p-value	ZU	Expected	p-value	W	ZW
LAIA	0.50	0.53	0.065	0.72	0.72	0.463	0.75	0.97
CARICOM	0.84	0.63	0.000	1.00	0.99	0.000	0.97	1.00
CAFTA-RD	0.76	0.77	0.366	0.62	0.62	0.542	0.92	0.93
SCM	0.83	0.85	0.211	0.88	0.88	0.500	0.59	0.92
CAIS	0.86	0.85	0.259	0.99	0.98	0.058	0.88	1.00
AC	0.88	0.89	0.345	0.97	0.97	0.464	0.86	0.99
NAFTA	0.93	0.93	0.548	0.86	0.86	0.676	-0.58	-0.64
OECS	1.00	0.93	0.006	1.00	1.00	0.008	1.00	1.00

Source: The Author (2021)

It is also analyzed the relational trends of regional groups in relation to trade agreements. The regional divisions are north, south and central. It is observed that both in the unweighted network and in the Z_unweighted network, the relations presented do not constitute a homophily propensity. The observed unweighted and Z_unweighted EI indexes do not differ significantly from the EI indexes obtained in a random network except the central region in the unweighted case, where trades with external countries are significantly higher than would be expected in a random network. Notwithstanding, when including the volume of transactions, the countries of the northern and central regions tend to trade more strongly toward countries in the same economic block. On the other hand, countries in the south region have a great increase in their tendency to trade with countries outside their trade agreements when trade volumes and countries' GDPs are simultaneously taken into account.

Table 7 – EI indexes for regional groups considering the trade agreements as class attributes in the Trade Network

Region	UW	Expected	p-value	ZU	Expected	p-value	W	ZW
South	0.43	0.42	0.305	0.71	0.71	0.530	0.03	0.90
North	0.52	0.52	0.525	0.46	0.46	0.518	-0.69	-0.72
Central	0.66	0.54	0.000	0.38	0.38	0.525	-0.18	-0.52

Source: The Author (2021)

5.4 CONCLUSIONS

In this chapter, it was proposed a new network measure which is a generalization of the EI index to measure homophily in cases where groups are non-disjoint. Two networks was explored with the new measure. In a co-authorship network, the non-disjoint groups were the areas of knowledge in which the researchers work. The analysis showed that operational research distinguished itself as the unique area having more internal relations, suggesting it to be a more independent research field. In a trade network, the non-disjoint groups were the blocks of trade agreements. In this network, surprisingly most countries have more external links. Despite that, this is not a trend, since the observed EI index does not differ significantly from the expected EI in a random network. Therefore, it is not possible to support that trade agreements might not be effectively helping the development of transaction between countries.

As it could be seen in these networks, the proposed measure allows the expansion of the social networks analysis. Through an homophily analysis, one can identify whether certain group of nodes have a tendency to work closely together or not. Currently, despite the wide possibility of finding non-disjoint groups, data is still scarce. However, it is hoped that the proposed measure will stimulate further studies and offers of data in online databases. It is also necessary to highlight that the calculation of the EI index occurred in four different situations: unweighted network (with and without the node's weight) and weighted network (with and without the node's weight). Leading to the enrichment of the analysis of the networks by indicating in which direction the relations are more intense or frequent and if they are influenced by the characteristics that define the nodes' weights.

6 A PROPOSAL FOR THE EI INDEX FOR FUZZY GROUPS

The results discussed in this chapter were submitted for publication in a journal and are under review.

6.1 CONTEXT

It is known that actors can belong to many associative groups simultaneously, with various levels of affiliation, and distinct disjoint groups rarely exist on a large scale in many empirical networks [36]. As mentioned in the previous chapter, currently publications that use the EI index as a measure of homophily are concentrated in disjoint or mutually exclusive groups. Situations where network actors are present in more than one group are not commonly explored. It is worth mentioning that one of the barriers encountered in the analysis of non-disjoint groups is the absence of a measure, since the EI index is defined for disjoint groups [107]. However, as presented in the previous chapter, Andrade and Rêgo [107] suggested a method that generalized the EI index developed by Krackhardt and Stern [57] and quantifies the relational structure within and between groups that encompass not only the analysis of disjoint groups but also of non-disjoint groups.

In this context, the objective of this chapter is to expand the generalized metric suggested by Andrade and Rêgo [107], adapting it to also encompass groups where the nodes present various levels of affiliations, fuzzy groups. Thereafter, it can be analyzed, for example, networks that analyze political behavior, studying relationships between voters with different positions in the political spectrum and networks of friendships with bilingual speakers, analyzing the relationships between speakers with different levels of language fluency. In this chapter, it is analyzed two networks. A co-authorship network formed by researchers with a PhD in production engineering, where the time of completion of the doctorate defined the fuzzy groups. The other network is formed by trade relations between American countries, we use the Human Development Index (HDI) to form fuzzy groups.

This chapter is organized as follows. In Section 6.2, it is presented the proposed measure, which is a generalization of the current EI index, encompassing fuzzy groups. Two applications of the proposed measure are made in Section 6.3. Finally, it is discussed the results of the applications in Section 6.4 and presented conclusions.

6.2 EI INDEX: FUZZY CASE

Every day, when describing certain phenomena (characteristics), it is used degrees that represent qualities or partial truths. As an example, the elderly group will be defined. A suggestion to formalize this set mathematically could have at least two approaches. The first, distinguishing from which age the individual is considered elderly. In this case, the set is well-defined. The second, less conventional, is given in such a way that individuals are considered elderly with more or less intensity, that is, there are elements that would belong more to the elderly class than others. This means that the younger the individual, the lower his or her degree of belonging to this class. Thus, it can be said that individuals belong to the class of elderly people, with more or less intensity. Mathematically, the sets to which the elements have degrees of pertinence are called fuzzy sets. The formalization of fuzzy sets was presented by Zadeh [114] as an extension of the classic notion of sets.

In order to explore cases of fuzzy groups, to obtain the EI index a new metric was developed, which is an adaptation of the metric proposed by Andrade and Rêgo [107] to generalize the original EI index measure for use with overlapping groups.

Let A be the set of all attributes for nodes in a social network with n nodes. For $X \in A$, let $\mu_X(v_i)$ be the level of membership of node v_i to a given group, $0 \leq \mu_X(v_i) \leq 1$. Moreover, for a generic set of nodes, S, consider the following sets of indices $I(S) = \{i : v_i \in S\}$ and $J^i(S) = \{j : (v_j \in S \text{ and } j > i) \text{ or } (v_j \notin S \text{ and } j \neq i)\}$. Thus, the number of external and internal links for a generic set of nodes, S, are given, respectively, by:

$$EL(S) = \sum_{i \in I(S)} \sum_{j \in J^i(S)} x_{ij}(1 - \max_{X \in A}\{\mu_X(v_i)\mu_X(v_j)\})$$

and

$$IL(S) = \sum_{i \in I(S)} \sum_{j \in J^i(S)} x_{ij} \max_{X \in A}\{\mu_X(v_i)\mu_X(v_j)\},$$

where in the unweighted case x_{ij} is either 1 or 0 depending on whether there is a link between node v_i and v_j, in the case with only edges' weights $x_{ij} = w_{ij}$ and in the case of edges' and nodes' weights $x_{ij} = z_{ij}$.

Alternatively, for $X \in A$, the number of external and internal links can be defined for the node group, S_X, that has attribute X, respectively, as follows:

$$EL(S_X) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}\mu_X(v_i)(1-\mu_X(v_j))$$

and

$$IL(S_X) = \sum_{i=1}^{n}\sum_{j>i} x_{ij}\mu_X(v_i)\mu_X(v_j),$$

where x_{ij} is defined exactly as before.

Since membership functions by definition assume values between 0 and 1 and the definitions of external and internal links involve products of membership functions, in order to avoid overestimating the external links, we recommend the use of trapezoidal membership functions. In order to obtain the trapezoidal membership functions, we suggest performing the following steps:

(i) Determine the highest value before which the degree of membership is known to be null.

(ii) Determine the lowest value from which it is known for certain that the degree of membership is null.

(iii) Determine the lowest value with degree of membership 1.

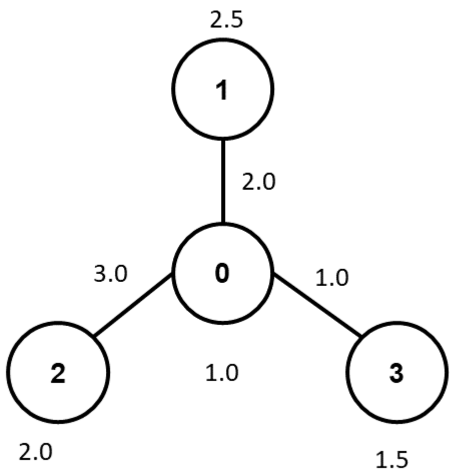
(iv) Determine the highest value with degree of membership 1.

To better explain the proposed method, here it is presented a simple example to expound how the new metric works in a specific network. Suppose there is a network with four nodes that belong with different membership levels to two groups, A and B (show in Figure 31). In the network, $set\{1,2\}$ is considered. Note that node 1 and 2 have no connections and that 0 is connected to both. Disregarding the edges' and nodes' weights, it is obtained $x_{10} = 1$ or $x_{01} = 1$, $\max_{X\in\{A,B\}}\{\mu_X(0)\mu_X(1), \mu_X(0)\mu_X(1)\} = \max_{X\in\{A,B\}}\{\ 0.65 * 0.80, 0.75 * 0.60\} = 0.52$ and $x_{20} = 1$ or $x_{02} = 1$, $\max_{X\in\{A,B\}}\{\mu_X(0)\mu_X(2), \mu_X(0)\mu_X(2)\} = \max_{X\in\{A,B\}}\{\ 0.65 * 0.65, 0.75 * 0.15\} = 0.4225$, so $EL(set\{1,2\}) = (1-0.52) + (1-0.4225) = 1.06$ and $IL(set\{1,2\}) = 0.52 + 0.4225 = 0.94$ and therefore, $EI(set\{1,2\}) = \frac{1.06-0.94}{1.06+0.94} = 0.06$. Now consider group A, we have the following edges x_{01} or x_{10}, x_{02} or x_{20} and x_{03} or x_{30}. $EL(groupA) = x_{01}\mu_X(0)(1-\mu_X(1)) + x_{10}\mu_X(1)(1-\mu_X(0)) + x_{02}\mu_X(0)(1-\mu_X(2)) + x_{20}\mu_X(2)(1-\mu_X(0)) + x_{03}\mu_X(0)(1-\mu_X(3)) + x_{30}\mu_X(3)(1-\mu_X(0)) = 0.65(1-0.80) +$

$0.80(1 - 0.65) + 0.65(1 - 0.65) + 0.65(1 - 0.65) + 0.65(1 - 0.25) + 0.25(1 - 0.65) = 0.13 + 0.28 + 0.2275 + 0.2275 + 0.4875 + 0.0875 = 1.44$ and $IL(groupB) = x_{01}\mu_X(0)\mu_X(1) + x_{02}\mu_X(0)\mu_X(2) + x_{03}\mu_X(0)\mu_X(3) = 0.65 * 0.80 + 0.65 * 0.65 + 0.65 * 0.25 = 1.105$, so $EI(groupA) = \frac{1.44-1.104}{1.44+1.104} = 0.13$.

Table 8 displays the results for the graph shown in Figure 31. It is easy to verify that the proposed metric is a generalization of the EI index proposed in [57] in the sense that if groups are disjoint and membership functions are either 0 or 1, then it coincides with (17).

Figure 31 – Social network with fuzzy groups of nodes.



	A	B
0	0.65	0.75
1	0.80	0.60
2	0.65	0.15
3	0.25	0.45

Source: The Author (2021)

Table 8 – EI Index fuzzy groups example

	Unweighted			Weighted		
	EL	IL	EI Index	EL	IL	EI Index
set {0}	1.72	1.28	0.15	3.36	2.65	0.12
set {1}	0.48	0.52	-0.04	0.96	1.04	-0.04
set {2}	0.58	0.42	0.16	1.73	1.27	0.16
set {3}	0.66	0.34	0.32	0.66	0.34	0.33
set {C}[a]	1.72	1.28	0.15	3.36	2.65	0.12
set {1,2}	1.06	0.94	0.06	2.69	2.31	0.08
set {1,3}	1.14	0.86	0.14	1.62	1.38	0.08
set {2,3}	1.24	0.76	0.24	2.40	1.61	0.20
Group A	1.44	1.10	0.13	2.76	2.47	0.06
Group B	1.65	0.90	0.29	3.45	1.58	0.37
	Z_unweighted			Z_weighted		
	EL	IL	EI Index	EL	IL	EI Index
set {0}	2.53	1.97	0.13	5.11	4.14	0.10
set {1}	0.84	091	-0.04	1.68	1.82	-0.04
set {2}	0.87	0.63	0.16	2.60	1.90	0.16
set {3}	0.83	0.42	0.33	0.83	0.42	0.33
set {C}	2.53	1.97	0.13	5.11	4.14	0.10
set {1,2}	1.71	1.54	0.05	4.28	3.72	0.07
set {1,3}	1.67	1.33	0.11	2.51	2.24	0.06
set {2,3}	1.69	1.06	0.23	3.43	2.32	0.19
Group A	2.12	1.75	0.10	4.20	3.92	0.03
Group B	2.33	1.38	0.26	5.14	2.50	0.34

[a] C is any set containing node 0 or the set {1,2,3}.

Source: The Author (2021)

6.3 APPLICATION

In this section, the method proposed is applied in two networks object of studies in previous publications. These networks present the fundamental element for our approach, which is the presence of fuzzy groups, in addition to information about the nodes' weights. As a means of comparison, it is also analyzed the cases of disjoint [34] and non-disjoint [107] groups. In this way, the EI Index will be obtained for 4 situations: without considering the weight of the edges and the nodes, unweighted (UW); regarding only the nodes' weight, Z_unweighted (ZU); considering only the edges' weight, weighted (W); taking into account both weights, Z_weighted (ZW).

To evaluate whether the EI Index for a given group is compatible with the expected when connections occur randomly, i.e., without preference of members for external or internal relations, for the unweighted and the Z_unweighted cases, the expected EI index is calculated for each one of the cases analyzed considering the average of 5000 randomly generated binomial graphs with the same density and size as that of the original graphs. It is also added a probability, p-value, that expresses how unlikely it is to obtain an EI index at least as extreme as the one observed in the randomly generated binomial graphs. One-sided p-values was considered, so that it is given by the relative frequency of times that the simulated EI has a value greater (resp., less) than or equal to the observed EI, when the expected EI is less (resp., greater) than the observed one.

6.3.1 Data

In order to implement the proposed EI index, data from two real networks are used. Next, it is given some details about these networks.

(i) Co-authorship PQ: The PQ network is a co-authorship network among researchers in the area of Production Engineering in Brazil that held a Productivity Research scholarship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) in 2015. It has 124 nodes and 131 edges. The network is undirected and the edges' weights represent the number of publications made in co-authorship by a given pair of researchers in the period of 2005 to 2014 [71].

(ii) Trade of American Countries: The network of commerce between American countries is formed by 30 countries and 356 edges. This network was created from the network of international trade developed by Andrade and Rêgo [14] which includes 178 countries from all continents, forming a unique main component with 10,419 edges. The network is undirected and the edges' weights represent the mean of export and import commercial transactions between a pair of countries during 2015.

6.3.2 PQ Network

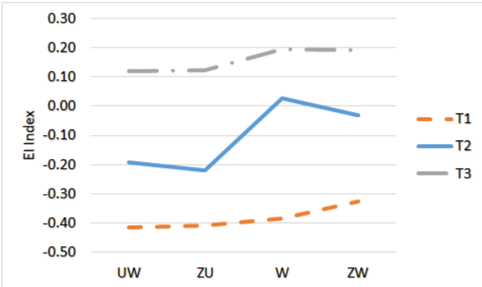
First, it is shown how the arbitrary choice of disjoint groups, according to the doctoral completion time, affects the EI index of these groups. Three cases of the disjoint groups (T1, T2 and T3) are delimited by varying the limits of the groups, Table 9, in the fuzzy regions, Table 10. Figure 32 shows the EI index for the entire network, for each of the arbitrary limits. It is clear that the result is strongly dependent on these limits.

Table 9 – Criteria for defining disjoint groups in the PQ Network

Case	PhD time	criterion	group size
T1	Young	≤ 5	2
	experient	≥ 6 and ≤ 24	100
	Senior	≥ 25	22
T2	Young	≤ 6	9
	experient	≥ 7 and ≤ 23	89
	Senior	≥ 24	26
T3	Young	≤ 7	12
	experient	≥ 8 and ≤ 20	73
	Senior	≥ 21	39

Source: The Author (2021)

Figure 32 – EI indexes for the whole PQ Network



Source: The Author (2021)

The definitions of the groups formed according to the doctoral completion time for the disjoint, non-disjoint and fuzzy case, followed the criteria in Table 10. For the disjoint case, the intermediary case T2 was considered.

Table 10 – Criteria for defining groups in the PQ Network

Case	PhD time	criterion	group size
Disjoint	Young	≤ 6	9
	experient	≥ 7 and ≤ 23	89
	Senior	≥ 24	26
Non-disjoint	Young	≤ 7	12
	experient	≥ 5 and ≤ 25	104
	Senior	≥ 20	45
Fuzzy	Young	$\mu(x) = \begin{cases} 1, & x < 5 \\ \frac{7-x}{2}, & 5 \leq x < 7 \end{cases}$	7 2
	experient	$\mu(x) = \begin{cases} \frac{x-5}{2}, & 5 < x \leq 7 \\ 1, & 7 < x < 20 \\ \frac{25-x}{5}, & 20 \leq x < 25 \end{cases}$	10 67 23
	Senior	$\mu(x) = \begin{cases} \frac{x-20}{5}, & 20 < x \leq 25 \\ 1, & x > 25 \end{cases}$	21 18

Source: The Author (2021)

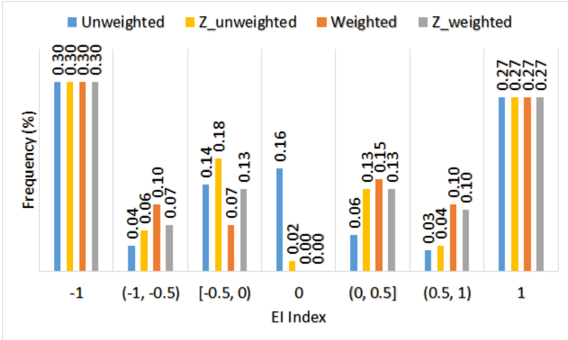
The researchers' h-index is used as the nodes' weights. The h-index is a measure that combines, in a simple way, the quantity of publications and the impact of publications and is given by the maximum value of h such that a researcher has published h works and each one of these works have been cited h or more times [112].

Figure 33 shows how the relationships between researchers occur. In general, most nodes in the non-disjoint case have an EI index of -1 (60%). In the fuzzy case and the disjoint case, the nodes have a similarity in relation to the proportion of EI index higher and lower than zero, however, in the fuzzy case the distribution of the EI index is more uniform.

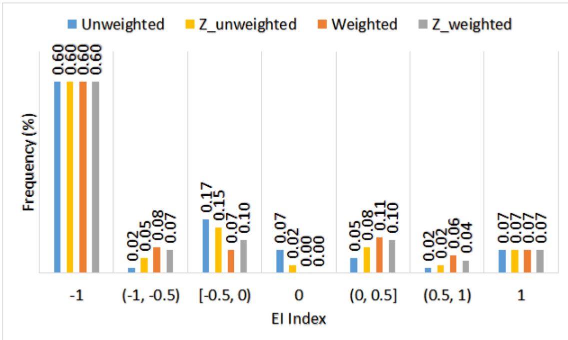
Figure 34 shows the EI index for the entire network. In general, when the nodes belong to non-disjoint groups, it is observed that the EI indexes are smaller, with a predominance of in-group relationships. On the other hand, when the groups are disjoint, the network has higher, but still negative, EI indexes. This means that, on a global level, there is a high level of cooperation between researchers from the same time group. As for the strength of the connections, it is observed that in the W network there is the lowest EI indexes and the ZW network the highest EI indexes. The first result indicates that the relationships are stronger between researchers in the same group and the second indicates that researchers who connect to researchers in other groups tend to link to researchers with higher h-indexes. It is worth mentioning that the negative EI indexes of the network, revealing a predominant in-group relationship, does not differ significantly from the result obtained in a random simulated network since the p-values are all greater than 0.05.

Figure 33 – Distribution of the EI index values in the PQ Network.

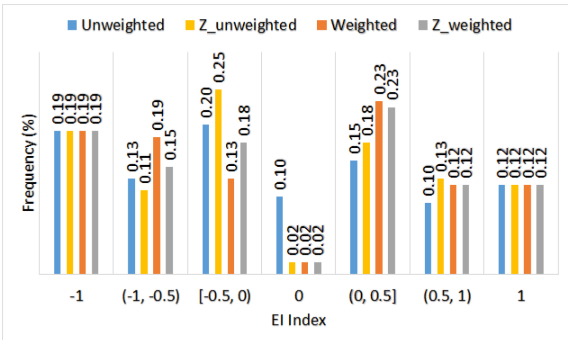
(a) Disjoint



(b) Non-disjoint

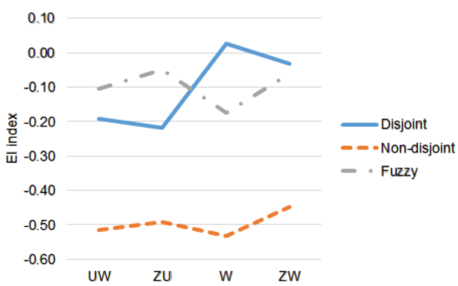


(c) Fuzzy



Source: The Author (2021)

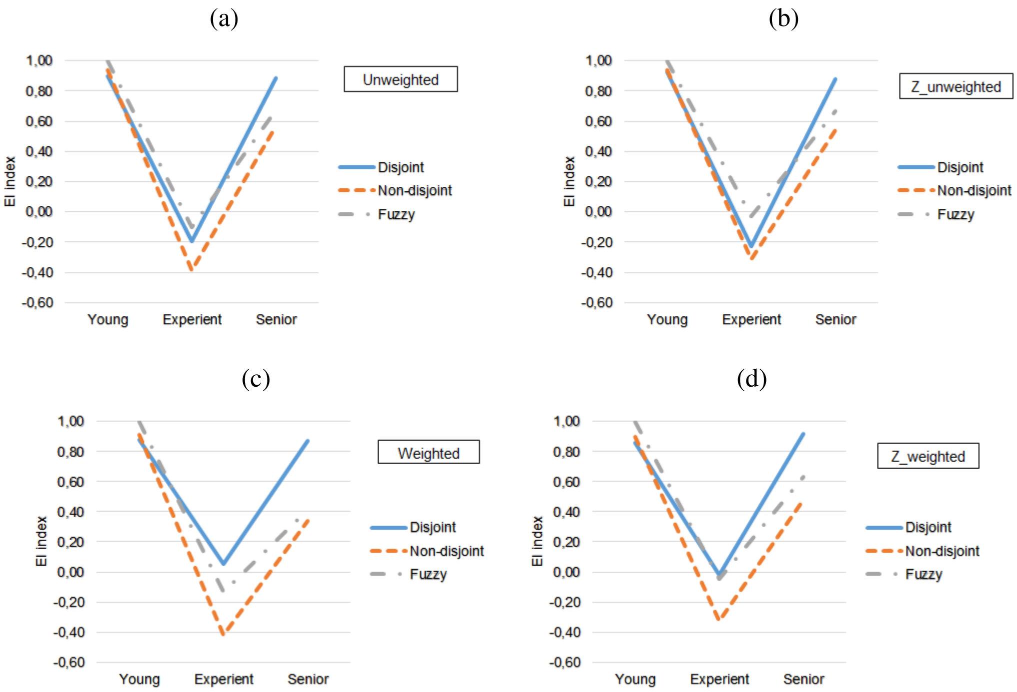
Figure 34 – EI indexes for the whole PQ Network



Source: The Author (2021)

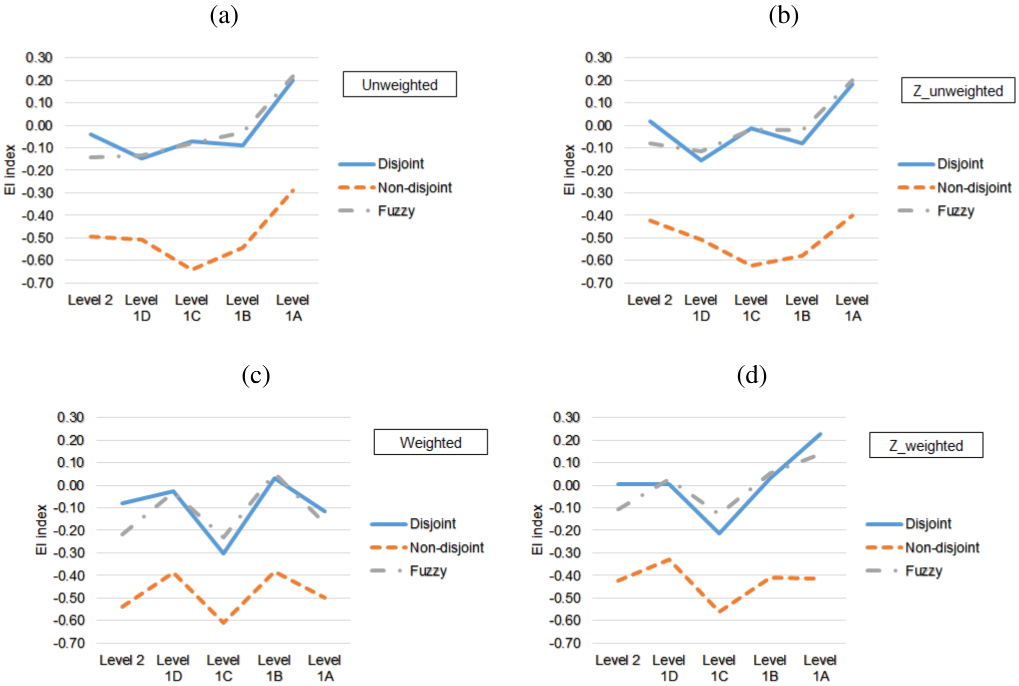
The analysis of the EI index of the experience level groups is shown in Figure 35. In general, when nodes belong to non-disjoint groups, it is observed that the EI indexes are smaller. In the case of disjoint and fuzzy groups, the EI indexes are close, with the EI indexes of the disjoint case slightly higher. The EI indexes of the experient group are negative, specially in the non-disjoint case. This shows that the internal connections of this group are higher than the external ones. The youth and senior groups have a positive EI index, with the youth being superior to the senior. This shows that the external relations of these outweigh the internal ones. Therefore, we can conclude that the experts cooperate with each other while young and senior PhDs are more conducive to cooperating with other groups. It is worth mentioning that the EI indexes obtained do not reveal a tendency towards homophily or heterophily, as they do not differ significantly from the results obtained by the random simulated network since the p-values are all greater than 0.05. Note that the weighting of the edges affected the EI index of the disjoint case more, making the relationships more heterogeneous. This is more noticeable in the case of experient groups.

Figure 35 – EI indexes for experience level areas in the PQ Network.



Source: The Author (2021)

Figure 36 – EI indexes for scholarship level groups considering the experience level as class attributes in the PQ Network.



Source: The Author (2021)

It was also analyzed the behavior of groups of researchers with the same level of scholarship in respect to the experience level group attributes. The scholarship level in order of importance and the total number of researchers are: 1A (8 %), 1B (5 %), 1C (8 %), 1D (19 %) and 2 (59 %). The analyses of the EI index of these groups are shown in Figure 36 for the cases of disjoint, non-disjoint and fuzzy groups, and studying the UW, ZU, W and ZW networks. In general, when the nodes belong to non-disjoint groups, it is observed that the EI indexes are smaller, predominance of in-group relationships. On the other hand, when the groups are disjoint or fuzzy, the network has higher EI indexes.

As for scholarship levels, there is a different behavior of the EI indexes for the different connection types, weighted or unweighted. Level 1A presents the highest EI indexes in the unweighted network, without or with the inclusion of the nodes' weights and in the weighted network considering the nodes' weights. Level 1A, the highest level of the scholarship, concentrates the most productive and influential researchers in the research area, being composed of 10 exclusively senior researchers and 2 exclusively experient. Although most are seniors, the in-group relationship is predominant in the non-disjointed case and external relationships are more common when the group is fuzzy or disjoint. Level 1A EI indexes are all negative in the weighted network. Level 1C, an intermediate scholarship level, also does not comprise young researchers. In the weighted network, without and with nodes' weights, as well as in the unweighted network (only in the non-disjoint case), the EI index of the level 1C is the smallest and negative. Therefore, for researchers at this level, most connections occur between researchers of the same experience level group. It is noteworthy that the EI indexes obtained do not reveal a tendency towards homophily or heterophily, as they do not differ significantly from the results obtained by random simultaed networks since the p-values are all greater than 0.05.

6.3.3 Trade of American Countries Network

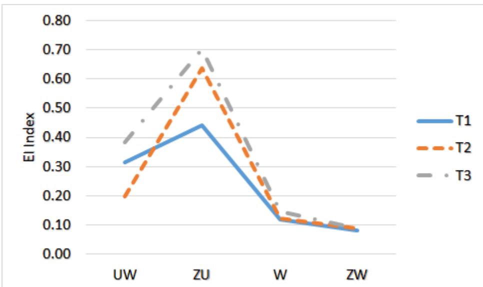
The Human Development Index (HDI) is used to form groups and first show how the arbitrary choice of disjoint groups, according to the HDI, affects the EI index of these groups. Three cases of the disjoint groups (T1, T2 and T3) are delimited by varying the limits of the groups, Table 11, in the fuzzy regions, Table 12. Figure 37 shows the EI index for the entire network, for each of the arbitrary limits. It is clear that the result is strongly dependent on these limits.

Table 11 – Criteria for defining groups in the Trade network

Case	HDI	criterion	group size
	Low	≤ 0.5	2
T1	Medium	> 0.5 and ≥ 0.675	5
	High	> 0.675 and ≥ 0.775	15
	Very high	> 0.775	8
	Low	≤ 0.5625	2
T2	Medium	> 0.5625 and ≥ 0.7	6
	High	> 0.7 and ≥ 0.8	18
	Very high	> 0.5	4
	Low	≤ 0.6	2
T3	Medium	> 0.6 and ≥ 0.725	12
	High	> 0.725 and ≥ 0.825	12
	Very high	> 0.825	4

Source: The Author (2021)

Figure 37 – EI indexes for the whole Trade Network



Source: The Author (2021)

The definitions of the groups formed according to the HDI for the disjoint, non-disjoint and fuzzy case, followed the criteria in the Table 12, where the intermediary case T2 was considered for the disjoint case.

Table 12 – Criteria for defining groups in the Trade Network

Case	HDI	criterion	group size
Disjoint	Low	≤ 0.5625	2
	Medium	> 0.5625 and ≤ 0.7	6
	High	> 0.7 and ≤ 0.8	18
	Very high	> 0.8	4
Non-disjoint	Low	≤ 0.6	2
	Medium	≥ 0.525 and ≤ 0.725	12
	High	≥ 0.675 and ≤ 0.825	19
	Very high	≥ 0.775	8
Fuzzy	Low	$\mu(x) = \begin{cases} 1, & x < 0.4 \\ \frac{0.6-x}{0.6-0.4}, & 0.4 \leq x < 0.6 \end{cases}$	0 2
	Medium	$\mu(x) = \begin{cases} \frac{x-0.525}{0.6-0.525}, & 0.525 < x \leq 0.6 \\ 1, & 0.6 < x < 0.675 \\ \frac{0.675-x}{0.675-0.725}, & 0.675 < x \leq 0.725 \end{cases}$	0 5 7
	High	$\mu(x) = \begin{cases} \frac{x-0.675}{0.6-0.725}, & 0.625 < x \leq 0.725 \\ 1, & 0.725 < x < 0.775 \\ \frac{0.825-x}{0.825-0.775}, & 0.775 \leq x < 0.825 \end{cases}$	11 8 4
	Very high	$\mu(x) = \begin{cases} \frac{x-0.775}{0.9-0.775}, & 0.775 < x \leq 0.9 \\ 1, & x > 0.9 \end{cases}$	6 2

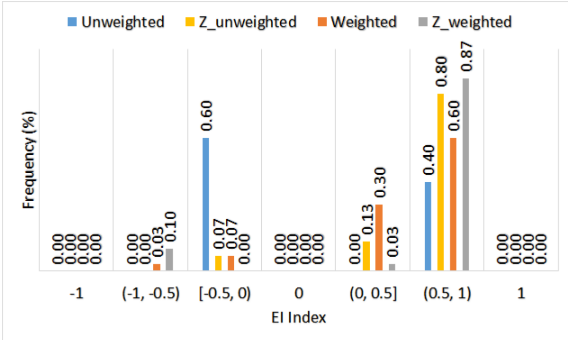
Source: The Author (2021)

Figure 38 shows the EI index at the individual level of the 30 countries. In general, countries have positive EI indexes, that is, intergroup relations superior to in-groups ones. In the non-disjoint case, it is possible to notice that some countries predominate in-group relations. The in-group relationship is also more visible when the network is unweighted.

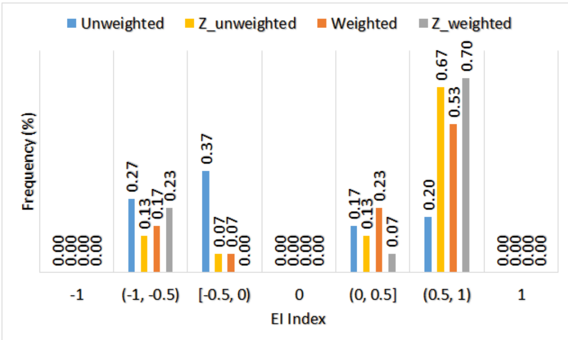
Figure 39 shows the EI index for the entire network. In general, when nodes belong to non-disjoint groups, it is observed that the EI indexes are smaller. On the other hand, when the groups are fuzzy, the network has higher EI indexes. The EI indexes are positive, except for the EI index in the case of non-disjoint groups in the weighted network. This indicates that, on a global level, trade occurs between countries of different HDI groups. The predominant intergroup relationships do not differ significantly from the result obtained by random simulated networks since the p-values are greater than 0.05.

Figure 38 – Distribution of EI index values in Trade Network.

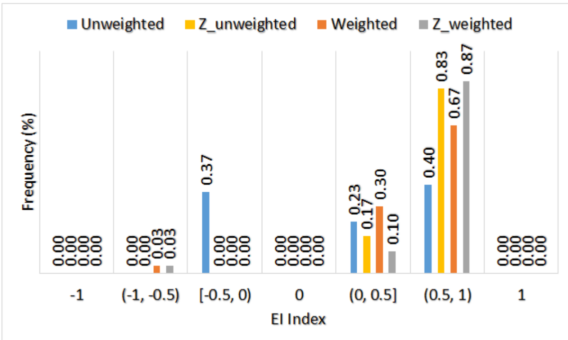
(a) Disjoint



(b) Non-disjoint

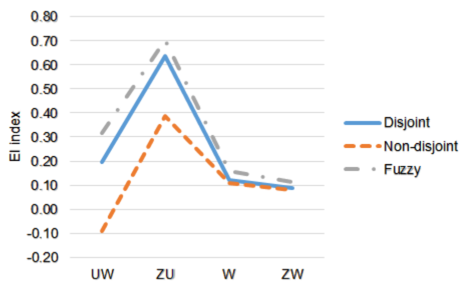


(c) Fuzzy



Source: The Author (2021)

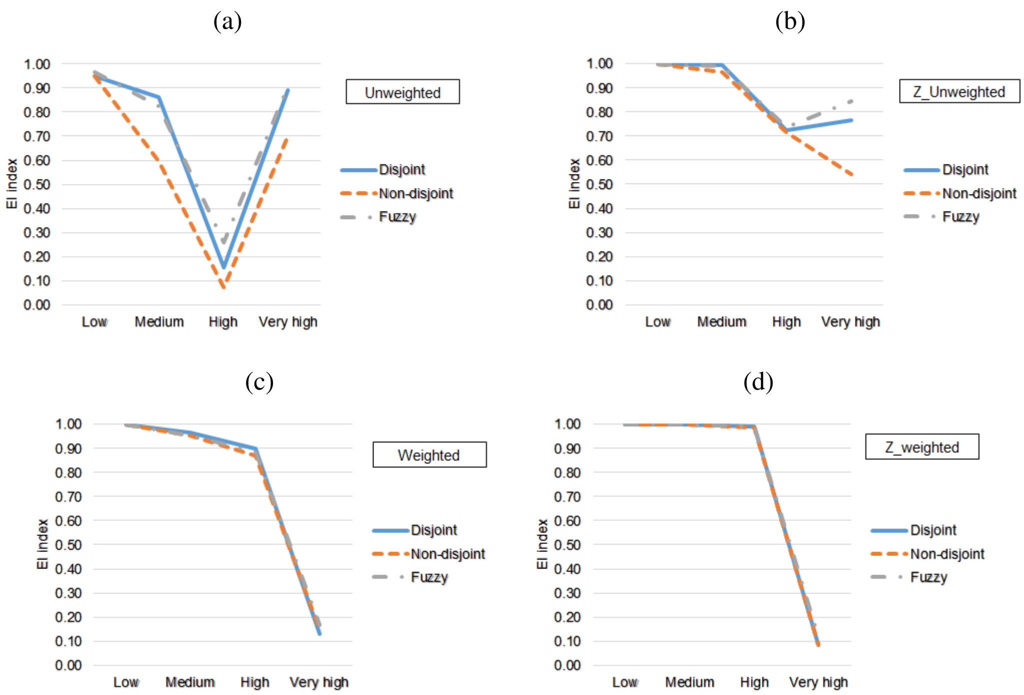
Figure 39 – EI indexes for the whole Trade Network



Source: The Author (2021)

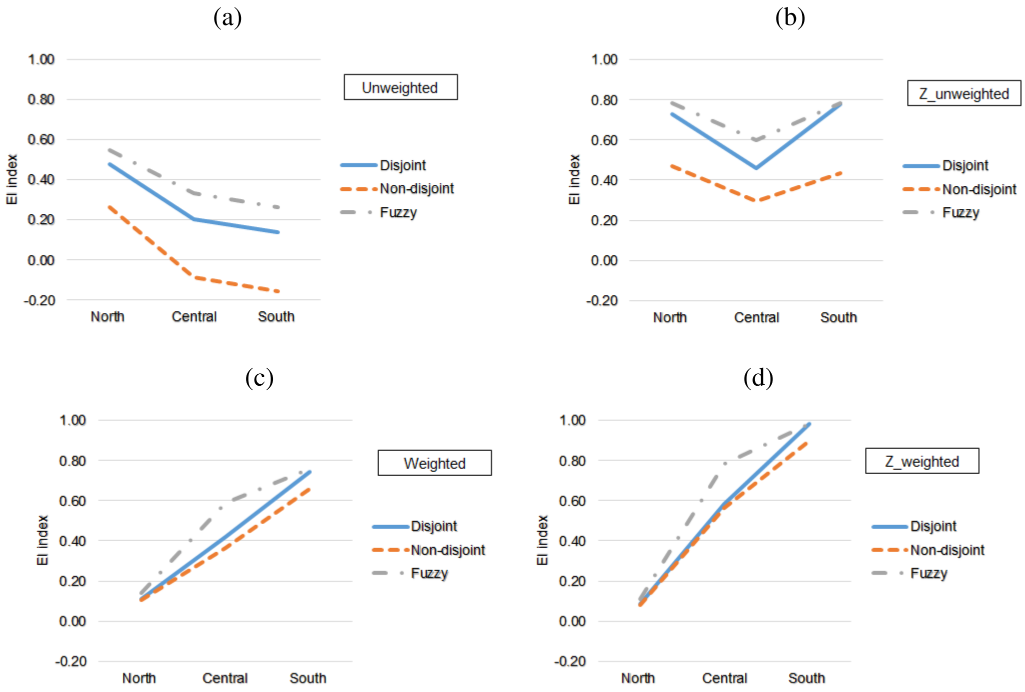
The analysis of the EI index of the HDI groups is shown in Figure 40. In general, the low and medium groups have the highest EI indexes, close to 1, countries in these groups have intergroup relations higher than in-groups, the EI indexes are statistically significant, that is, these groups are prone to heterophily. The group with high HDI has the lowest EI indexes in the unweighted network, being the one with the highest in-group relationship, but the EI indexes increases significantly in the Z_Unweighted, weighted and Z_Weighted network, that is, the relationships are stronger with other groups. The group of countries with a very high HDI has the lowest EI indexes in the weighted network, with and without the nodes' weights, revealing a stricter relationship between countries in the group. The EI indexes of the groups with high and very high HDI do not differ statistically from that presented by the random network.

Figure 40 – EI indexes for HDI groups in the Trade Network



Source: The Author (2021)

Figure 41 – EI indexes for regional groups considering the HDI groups as class attributes in the Trade Network



Source: The Author (2021)

It was also analyzed the behavior of groups of countries by region in respect to the HDI group attributes. The regional divisions are north, south and central, with 3, 12 and 15 countries, respectively. The analyses of the EI indexes of these groups are shown in Figure 41 for the cases of disjoint, non-disjoint and fuzzy groups, and studying the UW, ZU, W and ZW networks. In general, when the nodes belong to non-disjoint groups, it is observed that the EI indexes are smaller. On the other hand, when the groups are disjoint or fuzzy, the regions have higher EI indexes. As for the regions, there is a behavior different from the EI index depending on the connection type, weighted or unweighted. The northern region has the highest EI indexes on the UW and ZU networks. The northern region's EI indexes fall in the weighted network, indicating that countries in the northern region have stronger relations with countries in the same HDI group. The southern region in the UW network has the lowest EI indexes, positive in the disjoint and fuzzy case, and negative in the non-disjoint case. In weighted networks, with and without nodes' weights, the EI indexes are positive and higher in the southern region, indicating that the forces of relations are more intense between countries of different HDI groups. The EI indexes of the regions do not reveal a tendency towards homophily or heterophily, as they do not differ significantly from the EI presented in the simulated network with random relationships.

6.4 CONCLUSIONS

In this chapter, it was proposed a new network measure which is a generalization of the EI index to measure homophily in fuzzy group cases. Two networks was explored with the new measure. In a co-authorship network, the doctoral completion time was used to form groups. In a trade network among countries, the Human Development Index (HDI) was used to form groups. It was obtained the EI index for the networks considering the cases of disjoint, non-disjoint and fuzzy groups, and analyzing different relational forces, unweighted, weighted, without and with the nodes' weights. The proposed measure, applied in these two networks, allows for an expansion of the analysis of social networks. Through a homophily analysis, it is possible to identify whether a certain group of nodes has a tendency to work together or not.

In general, it is clear that fuzzy groups generate more homogeneous cooperation or commercial relations. This was already expected due to the fact that the actors present multiple associations with the same degree of association, equal to 1. In the co-authorship network, it was noticed that the researchers allocated as experient are the ones that most cooperate with each other. These relationships are favored because there are more experient researchers. The smaller number of young and senior researchers also justifies the predominance of external relations by these researchers. In the trade network, it was realized that relations between countries with different levels of development are more common. In the case of the groups with low and medium HDI, it was noticed that the EI index close to 1 is statistically significant, revealing the tendency towards heterophily in these two groups, revealing their dependency on more developed countries.

7 FINAL CONSIDERATIONS

In this final chapter, we summarize the main contributions of this work, discuss some of its limitations and propose directions for future works in the field.

7.1 CONTRIBUTIONS

SNA as a field of study aims at investigating the structure of relationships that connect individuals (or other social units, such as organizations) and how these relationships influence behavior and attitudes. This dissertation made four contributions to the advancement of studies using networks and graph theory to investigate the social structure. Generally speaking, all of these contributions were in the proposal of new metrics for SNA, motivated by an absence or deficiency in the literature. Properties of the proposed metrics were explored and applied to real social network to demonstrate their usefulness. Thus, the specific objectives established in the introduction of this work were effectively achieved.

The metrics proposed in this work helped to better identify, for example, how countries are positioned in the global trade network and how their commercial relations are influenced by trade agreements or by the Human Development Index of their partners. Therefore, it is possible to affirm that this work draw in social and economic impacts, as this understanding can help in the strategic planning of countries regarding their international trades. This work also has the potential of helping in the combat of criminal organizations, as it proposes a metric to target more efficiently individuals in a network of encounters of individuals using anklet monitors. Thus, the social impact of this work is evident, since, such information can be used by police forces to direct their efforts in the monitoring of the most important suspects, helping in criminality reduction. Finally, this work also has an impact on the academic community of Production Engineering in Brazil since it analyzed a coauthoship network of researchers fellows from CNPq in the area of Production Engineering revealing the most influential researchers and how coauthorship are influenced by the area of work of the researcher or by their scholarship levels. In addition to the social and economic impacts presented, given in the applications, the metrics proposed in the work may have environmental impacts. For example, it can contribute to studies such as Barnes et al. [115], that found that social networks (among fishermen) “are tied to actions that can directly impact marine ecosystems, and that biases toward within-group ties may impede the diffusion of sustainable behaviors.”

Next, it is highlighted the main conclusions of each proposed measure.

(i) Chapter 3 - inspired by the similarity between some geometric centrality metrics and the generalized mean of numbers, a new measure of centrality is proposed, which was called p-means centrality. The main characteristic of this measure is the ability to move among different measures of nodes' importance, only changing the p-parameter. The p-means centrality includes as special cases the well-known degree, eccentricity, closeness and harmonic centralities. To evaluate the performance of the p-means centrality, it is implemented on four real networks and compared to the spread capacity simulated using a SIR model. It was employed the Kendall's tau coefficient τ to measure the rank correlation between the ranking list generated by the simulation results and ranking lists generated by the p-means centrality for different values of p. The p-parameters were chosen to maximize the number of different classification of the nodes and the simulation of the SIR model was performed for different values of probability of infection, also choosing the values which maximize the number of different classification of the nodes. It was also evaluated the Kendall correlation between the p-means centrality and other centrality measures: degree centrality, betweenness centrality, eigenvector centrality and PageRank. The p-means centrality is important because it allows exploring the effect of distance variation on the influence of nodes, classifying nodes with greater spread capacity than known geometric measures and other traditional ones.

(ii) Chapter 4 - it was proposed a new measure of centrality based on the gravity law, called Energy Disruptive Centrality (ED), in order to identify nodes that, when removed from the network, lead to its disruption, structurally and functionally. As a new approach, this study also addresses how the anklet network was formed. The network, in addition to meeting the objective of this study, will also be useful for future works that may address the relational patterns and socio-economic characteristics of those of the co-offenders and thus be able to help in the understanding and analysis of crime patterns with greater precision and, consequently, improve crime prevention.

(iii) Chapter 5 and Chapter 6 - a new network measure which is a generalization of the EI index to measure homophily in cases where groups are non-disjoint was proposed. And as an extension of this work, a new network measure which is a generalization of the EI index to measure homophily in cases where groups are fuzzy was proposed. We highlight that the only work found that mentions the use of the EI index for non-disjoint groups was that of [34], but that it did not clarify all the issues involved in the problem. Given that disjoint groups on a large scale rarely exist in many empirical networks, analysis of non-disjoint and fuzzy groups is relevant, and the proposed measure makes it possible to explore these networks. Therefore, this work fills this gap in the literature. Although non-disjoint and fuzzy groups are more abundant, data are scarce for network analysis. In what regards the EI index for overlapping groups proposed, extraordinary situations that may occur may limit its use. For example, there may be a single attribute group that involves all or almost all nodes. In this case, the EI index for any set of nodes would be close to or equal to -1, since all or almost all edges would be classified as internal, regardless of the participation of the nodes in the other attribute groups. A problem that generates this limitation is that it was assumed that the participation of nodes in attribute groups is homogeneous, i.e., it is unable to capture the idea that a node may have more relationship with one attribute group than with another. To mitigate this problem, the EI index for fuzzy groups proposed may be used. In what regards the EI index for fuzzy groups, one must be very careful in what shapes of level of memberships are used, since if most nodes have low degree of membership to all groups, the EI index can be artificially higher than expected. We viewed that trapezoidal membership functions perform better than triangular ones, for example.

7.2 LIMITATIONS AND FUTURE WORKS

(i) The p-means centrality was designed for application in an unweighted network, it does not consider the edge's weigh. For future work, the measure can be adapted for application in weighted networks as well. Furthermore, it has not been explored as the possibility of placing the node's weights in this type of metric. A generalized weighted mean using node's attributes as weights can be tested. In this work, we analyzed the effectiveness of p-means centrality by the SIR transmission model, but the effectivity of the measure could be analyzed in another way, for example, by the ability to dismantle the network.

(ii) When applying the energy disruptive centrality in anklet network, we chose to only use as an attribute the anklet displacement information, so that it does not present a bias when making a comparison with other centrality measures in terms of detecting certain types of crimes. However, other nodes' attributes could be used. For example, if the objective is to apply the attribute gravity centrality to determine the most influential individuals in the network, one could adopt the number of years of conviction, whether the convict is a repeat offender or not, the seriousness of the committed crime, among others. Future work may address different measures of network damage. For example, [91] addresses six metrics that quantify the structural properties of networks and their abilities to disseminate information. It is worth pointing out that the ethical side of taking measures based on the role of individual in networks must be taken care of. So, how public policies can be made in light of the generated information from the network analysis? In the case of the convicts wearing anklet monitors, it is not the case that one should limit the freedom of specific individuals only on the basis of the network analysis results, but these results can be used by the policy in order to closer monitor the routine of such individuals in order to find out whether they are really committing crimes and, only in this case, to imprison them. Finally, it is important to emphasize that, as it is the case for any other centrality measure, it is not expected that the proposed metrics should be the best ones in all kinds of networks. Studying in what types of networks the proposed metrics behave optimally is an important field for future research.

(iii) It is worth pointing out that there is some limitation in the use of EI index measures to determine if links are indeed established due to the analyzed attributes. EI index measures indicate the degree of homogeneity or heterogeneity of the relationships between the groups involved, and this analysis takes place considering a certain attribute. As it was done in this work, these measures can be used, for example, to test whether more homogeneous or heterogeneous relations occur in a given network than in a random network. However, these measures by themselves do not take into account dynamical features of the networks and, consequently, cannot infer causal relationships. An important open question is whether a dynamical analysis of networks can use EI measures to determine whether nodes' attributes influence link formation, or if the association between nodes' attributes and links is only spurious. The analysis of dynamic networks is linked to temporal analysis, which can be observed that changes in networks can be the result of external factors that are independent of the social characteristics found in the networks. Therefore, by performing a dynamic analysis, we can reach more assertive conclusions about the variables that influence the formation of networks.

Given the promising results of the measures proposed in this work, it is expected to provide for future research in SNA new options for measures that seek to identify influential actors or spreaders in networks in different scopes of studies, as well as the analysis of network formations. The proposed measures stand out for exploring the identified gaps: the influence of the variation in distances between nodes; adding (defining) weights for nodes based on their attributes; and the relationships between non-disjoint or fuzzy groups in the formation of networks.

The co-authorship networks, sources of studies of several works, as well as in this one, present some challenges that are worth exploring. For example, How do you differentiate when there are too many authors in a work? Does the order of co-authorship imply the importance of the author? For the first, we can weight edges by the inverse of the number of co-authors, making the links weaker as the number of co-authors increases. For the last question, the solution could be a directed network, from the author with the highest participation to the one with the lowest. Another interesting approach is the comparison between co-authorship networks of publications in journals and events to investigate whether there is a difference in the form of interaction and propagation of information.

It should also be noted that although the proposed measures were applied in specific cases, co-authorship networks, international trade networks and criminal networks, the flexibility of this methodology has allowed it to be applied in different fields of research. As mentioned in the introduction, Chapter 1, industrial engineering has benefited from the use of this methodology in solving problems: operations planning [38], quality management [3] and project management [39]. The measures proposed in this work can also be applied in the future to solve these problems, among others.

References

1 NEWMAN, M. E. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, National Acad Sciences, v. 98, n. 2, p. 404–409, 2001. Citado 3 vezes nas páginas 17, 18, and 26.

2 NEWMAN, M. E. J. Scientific collaboration networks. i. network construction and fundamental results. Phys. Rev. E, American Physical Society, v. 64, p. 016131, Jun 2001. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevE.64.016131>. Citado na página 17.

3 SITKO-LUTEK, A. et al. Applying social network analysis on customer complaint handling. Industrial management & data systems, Emerald Group Publishing Limited, 2010. Citado 3 vezes nas páginas 17, 22, and 113.

4 WASSERMAN, S.; FAUST, K. Social network analysis: Methods and applications. [S.l.]: Cambridge university press, 1994. v. 8. Citado 3 vezes nas páginas 17, 28, and 53.

5 PHILLIPS, D. M.; PHILLIPS, J. K. A social network analysis of business logistics and transportation. International Journal of Physical Distribution & Logistics Management, MCB UP Ltd, v. 28, n. 5, p. 328–348, 1998. Citado na página 17.

6 GIRVAN, M.; NEWMAN, M. E. Community structure in social and biological networks. Proceedings of the national academy of sciences, National Acad Sciences, v. 99, n. 12, p. 7821–7826, 2002. Citado na página 17.

7 MAIREDER, A. et al. Big data and political social networks: Introducing audience diversity and communication connector bridging measures in social network theory. Social Science Computer Review, SAGE Publications Sage CA: Los Angeles, CA, v. 35, n. 1, p. 126–141, 2017. Citado na página 17.

8 WEY, T. et al. Social network analysis of animal behaviour: a promising tool for the study of sociality. Animal behaviour, Elsevier, v. 75, n. 2, p. 333–344, 2008. Citado na página 17.

9 HAMEDE, R. K. et al. Contact networks in a wild tasmanian devil (sarcophilus harrisii) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. Ecology letters, Wiley Online Library, v. 12, n. 11, p. 1147–1157, 2009. Citado na página 17.

10 ORTIZ-PELAEZ, A. et al. Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (fmd) epidemic in the uk. Preventive veterinary medicine, Elsevier, v. 76, n. 1, p. 40–55, 2006. Citado na página 17.

11 WANG, S.; DU, Y.; DENG, Y. A new measure of identifying influential nodes: Efficiency centrality. Communications in Nonlinear Science and Numerical Simulation, Elsevier, v. 47, p. 151–163, 2017. Citado 3 vezes nas páginas 17, 28, and 42.

12 BIAN, T.; HU, J.; DENG, Y. Identifying influential nodes in complex networks based on ahp. Physica A: Statistical Mechanics and its Applications, Elsevier, v. 479, p. 422–436, 2017. Citado 2 vezes nas páginas 17 and 42.

13 ANDRADE, R. L. de; RÊGO, L. C. p-means centrality. Communications in Nonlinear Science and Numerical Simulation, Elsevier, v. 68, p. 41–55, 2019. Citado 12 vezes nas páginas 17, 33, 37, 38, 42, 44, 45, 46, 47, 48, 49, and 57.

14 ANDRADE, R. L. de; RÊGO, L. C. The use of nodes attributes in social network analysis with an application to an international trade network. Physica A: Statistical Mechanics and its Applications, Elsevier, v. 491, p. 249–270, 2018. Citado 8 vezes nas páginas 17, 19, 30, 31, 32, 55, 85, and 97.

15 BRIGHT, D. A. et al. The use of actor-level attributes and centrality measures to identify key actors: A case study of an australian drug trafficking network. Journal of Contemporary Criminal Justice, SAGE Publications Sage CA: Los Angeles, CA, v. 31, n. 3, p. 262–278, 2015. Citado 6 vezes nas páginas 17, 18, 19, 52, 55, and 62.

16 SCHWARTZ, D. M.; ROUSELLE, T. D. Using social network analysis to target criminal networks. Trends in Organized Crime, Springer, v. 12, n. 2, p. 188–207, 2009. Citado 6 vezes nas páginas 17, 19, 52, 54, 55, and 68.

17 MA, L.-l. et al. Identifying influential spreaders in complex networks based on gravity formula. Physica A: Statistical Mechanics and its Applications, Elsevier, v. 451, p. 205–212, 2016. Citado 3 vezes nas páginas 17, 19, and 51.

18 WANG, J.; LI, C.; XIA, C. Improved centrality indicators to characterize the nodal spreading capability in complex networks. Applied Mathematics and Computation, Elsevier, v. 334, p. 388–400, 2018. Citado 3 vezes nas páginas 17, 19, and 51.

19 RYSZ, M.; PAJOUH, F. M.; PASILIAO, E. L. Finding clique clusters with the highest betweenness centrality. European Journal of Operational Research, Elsevier, v. 271, n. 1, p. 155–164, 2018. Citado na página 17.

20 NASIRIAN, F.; PAJOUH, F. M.; BALASUNDARAM, B. Detecting a most closeness-central clique in complex networks. European Journal of Operational Research, Elsevier, v. 283, n. 2, p. 461–475, 2020. Citado na página 17.

21 FREEMAN, L. C. Centrality in social networks conceptual clarification. Social Networks, Elsevier, v. 1, n. 3, p. 215–239, 1978. Citado 6 vezes nas páginas 17, 26, 27, 28, 29, and 53.

22 BONACICH, P. Power and centrality: A family of measures. American Journal of Sociology, University of Chicago Press, v. 92, n. 5, p. 1170–1182, 1987. Citado 3 vezes nas páginas 17, 29, and 53.

23 ROBINS, G. Understanding individual behaviors within covert networks: the interplay of individual qualities, psychological predispositions, and network effects. Trends in Organized Crime, Springer, v. 12, n. 2, p. 166–187, 2009. Citado 2 vezes nas páginas 17 and 18.

24 LIU, J. et al. A new method to construct co-author networks. Physica A: Statistical Mechanics and its Applications, Elsevier, v. 419, p. 29–39, 2015. Citado 5 vezes nas páginas 17, 19, 29, 30, and 55.

25 NEWMAN, M. E. Scientific collaboration networks. i. network construction and fundamental results. Physical review E, APS, v. 64, n. 1, p. 016131, 2001. Citado na página 18.

26 BOLDI, P.; VIGNA, S. Axioms for centrality. Internet Mathematics, Taylor & Francis, v. 10, n. 3-4, p. 222–262, 2014. Citado 8 vezes nas páginas 18, 27, 28, 33, 34, 38, 39, and 40.

27 WANDERLEY, A. J. et al. Identificando correlações entre métricas de análise de redes sociais e o h-index de pesquisadores de ciência da computação. In: SBC. Anais do III Brazilian Workshop on Social Network Analysis and Mining. [S.l.], 2014. p. 45–56. Citado na página 18.

28 LINDELAUF, R. H.; HAMERS, H. J.; HUSSLAGE, B. Cooperative game theoretic centrality analysis of terrorist networks: The cases of jemaah islamiyah and al qaeda. European Journal of Operational Research, Elsevier, v. 229, n. 1, p. 230–238, 2013. Citado 4 vezes nas páginas 19, 53, 73, and 74.

29 NIU, J.; YANG, H.; WANG, L. Logarithmic gravity centrality for identifying influential spreaders in dynamic large-scale social networks. In: IEEE. 2017 IEEE International Conference on Communications (ICC). [S.l.], 2017. p. 1–6. Citado 2 vezes nas páginas 19 and 51.

30 LI, Z. et al. Identifying influential spreaders by gravity model. Scientific reports, Nature Publishing Group, v. 9, n. 1, p. 1–7, 2019. Citado 2 vezes nas páginas 19 and 51.

31 SCHMID-PETRI, H. et al. Homophily and prestige: An assessment of their relative strength to explain link formation in the online climate change debate. Social networks, Elsevier, v. 55, p. 47–54, 2018. Citado na página 20.

32 MCPHERSON, M.; SMITH-LOVIN, L.; COOK, J. M. Birds of a feather: Homophily in social networks. Annual review of sociology, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 27, n. 1, p. 415–444, 2001. Citado 3 vezes nas páginas 20, 30, and 31.

33 YU, Z. et al. Structured subspace embedding on attributed networks. Information Sciences, Elsevier, v. 512, p. 726–740, 2020. Citado na página 20.

34 EVERETT, M. G.; BORGATTI, S. P. Categorical attribute based centrality: E-i and g-f centrality. Social Networks, Elsevier, v. 34, n. 4, p. 562–569, 2012. Citado 5 vezes nas páginas 20, 31, 81, 96, and 110.

35 KRACKHARDT, D. Graph theoretical dimensions of informal organizations. Computational organization theory, Lawrence Erlbaum and Associates Hillsdale, NJ, v. 89, n. 112, p. 123–140, 1994. Citado 2 vezes nas páginas 20 and 31.

36 LESKOVEC, J. et al. Statistical properties of community structure in large social and information networks. In: Proceedings of the 17th international conference on World Wide Web. [S.l.: s.n.], 2008. p. 695–704. Citado 3 vezes nas páginas 20, 81, and 92.

37 KONG, X. et al. Academic social networks: Modeling, analysis, mining and applications. Journal of Network and Computer Applications, Elsevier, v. 132, p. 86–103, 2019. Citado na página 21.

38 VARELA, M. et al. Using social network analysis for industrial plant layout analysis in the context of industry 4.0. International Journal of Industrial and Systems Engineering, Inderscience Publishers (IEL), v. 34, n. 1, p. 1–19, 2020. Citado 2 vezes nas páginas 22 and 113.

39 LI, Y.; LU, Y. Social network model of complex projects organization. In: IEEE. 2009 International Conference on Management and Service Science. [S.l.], 2009. p. 1–6. Citado 2 vezes nas páginas 22 and 113.

40 PIGATTO, G. et al. Competitiveness and social network of brazilian fish farmers. Journal of Agribusiness in Developing and Emerging Economies, Emerald Publishing Limited, 2020. Citado na página 22.

41 NEWMAN, M. E. Coauthorship networks and patterns of scientific collaboration. Proceedings of the national academy of sciences, National Acad Sciences, v. 101, n. suppl 1, p. 5200–5205, 2004. Citado 3 vezes nas páginas 25, 26, and 29.

42 BRANDES, U. A faster algorithm for betweenness centrality. Journal of mathematical sociology, Taylor & Francis, v. 25, n. 2, p. 163–177, 2001. Citado na página 26.

43 BARRAT, A. et al. The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America, National Acad Sciences, v. 101, n. 11, p. 3747–3752, 2004. Citado na página 26.

44 ABBASI, A.; HOSSAIN, L.; LEYDESDORFF, L. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. Journal of Informetrics, Elsevier, v. 6, n. 3, p. 403–412, 2012. Citado 2 vezes nas páginas 26 and 29.

45 LIU, X. et al. Co-authorship networks in the digital library research community. Information processing & management, Elsevier, v. 41, n. 6, p. 1462–1480, 2005. Citado na página 26.

46 HAGE, P.; HARARY, F. Eccentricity and centrality in networks. Social networks, Elsevier, v. 17, n. 1, p. 57–63, 1995. Citado na página 27.

47 MARCHIORI, M.; LATORA, V. Harmony in the small-world. Physica A: Statistical Mechanics and its Applications, Elsevier, v. 285, n. 3, p. 539–546, 2000. Citado 2 vezes nas páginas 28 and 33.

48 NEWMAN, M. E. The structure and function of complex networks. SIAM review, SIAM, v. 45, n. 2, p. 167–256, 2003. Citado na página 28.

49 ROCHAT, Y. Closeness centrality extended to unconnected graphs: The harmonic centrality index. In: ASNA. [S.l.: s.n.], 2009. p. 1–14. Citado 4 vezes nas páginas 28, 33, 37, and 57.

50 DEKKER, A. Conceptual distance in social network analysis. Journal of Social Structure (JOSS), v. 6, 2005. Citado 2 vezes nas páginas 28 and 57.

51 JACKSON, M. Social and Economic Networks, Princeton, NJ: Princeton Univ. [S.l.]: Press, 2008. Citado na página 29.

52 PAGE, L. et al. The PageRank citation ranking: Bringing order to the web. [S.l.], 1999. Citado na página 29.

53 BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems, Elsevier, v. 30, n. 1, p. 107–117, 1998. Citado 2 vezes nas páginas 29 and 53.

54 BURT, R. S. Measuring age as a structural concept. Social Networks, Elsevier, v. 13, n. 1, p. 1–34, 1991. Citado 3 vezes nas páginas 30, 31, and 80.

55 BENJUMEA, A. A. Homophily and Ethnic Background in the Classroom. 2015. Citado 2 vezes nas páginas 30 and 80.

56 LAZARSFELD, P. F.; MERTON, R. K. et al. Friendship as a social process: A substantive and methodological analysis. Freedom and control in modern society, New York, Van Nostrand, v. 18, n. 1, p. 18–66, 1954. Citado na página 30.

57 KRACKHARDT, D.; STERN, R. N. Informal networks and organizational crises: An experimental simulation. Social psychology quarterly, JSTOR, p. 123–140, 1988. Citado 5 vezes nas páginas 31, 80, 83, 92, and 95.

58 UCINET, U. Ucinet 6 for windows software for social network analysis. 1999. Citado na página 31.

59 HANNEMAN, R. A.; RIDDLE, M. Introduction to social network methods. [S.l.]: University of California Riverside, 2005. Citado na página 31.

60 BOJANOWSKI, M.; CORTEN, R. Measuring segregation in social networks. Social Networks, Elsevier, v. 39, p. 14–32, 2014. Citado na página 31.

61 SWEET, T. M.; ZHENG, Q. A mixed membership model-based measure for subgroup integration in social networks. Social Networks, Elsevier, v. 48, p. 169–180, 2017. Citado na página 31.

62 HENRY, A. D.; MITSCHE, D.; PRAŁAT, P. Homophily, influence and the decay of segregation in self-organizing networks. Network Science, Cambridge University Press, v. 4, n. 1, p. 81–116, 2016. Citado na página 31.

63 FREEMAN, L. C. Segregation in social networks. Sociological Methods & Research, Sage Publications Sage CA: Thousand Oaks, CA, v. 6, n. 4, p. 411–429, 1978. Citado na página 31.

64 DANCHEV, V.; PORTER, M. A. Heterogeneity of global and local connectivity in spatial network structures of world migration. Available at SSRN 2755271, 2016. Citado na página 31.

65 AGNEESSENS, F.; BORGATTI, S. P.; EVERETT, M. G. Geodesic based centrality: Unifying the local and the global. Social Networks, Elsevier, v. 49, p. 12–26, 2017. Citado na página 33.

66 KRACKHARDT, D. The strength of strong ties: The importance of philos in organizations. In Networks and Organizations: Structure, Form, and Action, Harvard Business School Press, p. 216–239, 1992. Citado na página 34.

67 KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences, The Royal Society, v. 115, n. 772, p. 700–721, 1927. Citado na página 34.

68 SHELDON, N. The generalized mean. Teaching Statistics, Wiley Online Library, v. 26, n. 1, p. 24–25, 2004. Citado na página 34.

69 BORWEIN, J. M.; BORWEIN, P. B. Pi and the AGM. [S.l.]: Wiley, New York, 1987. Citado na página 34.

70 NEWMAN, M. E. Finding community structure in networks using the eigenvectors of matrices. Physical review E, APS, v. 74, n. 3, p. 036104, 2006. Citado na página 41.

71 ANDRADE, R. L. d.; RÊGO, L. C. Exploring the co-authorship network among cnpq's productivity fellows in the area of industrial engineering. Pesquisa Operacional, SciELO Brasil, v. 37, n. 2, p. 277–310, 2017. Citado 3 vezes nas páginas 41, 85, and 97.

72 LÜ, L. et al. Leaders in social networks, the delicious case. PloS one, Public Library of Science, v. 6, n. 6, p. e21202, 2011. Citado na página 42.

73 ANDRADE, R. L. de et al. Energy disruptive centrality with an application to criminal network. Communications in Nonlinear Science and Numerical Simulation, Elsevier, v. 99, p. 105834, 2021. Citado 13 vezes nas páginas 51, 62, 63, 65, 66, 67, 70, 73, 74, 75, 76, 77, and 78.

74 GARAS, A.; SCHWEITZER, F.; HAVLIN, S. A k-shell decomposition method for weighted networks. New Journal of Physics, IOP Publishing, v. 14, n. 8, p. 083030, 2012. Citado na página 51.

75 KITSAK, M. et al. Identification of influential spreaders in complex networks. Nature physics, Nature Publishing Group, v. 6, n. 11, p. 888–893, 2010. Citado na página 51.

76 QI, X. et al. Terrorist networks, network energy and node removal: A new measure of centrality based on laplacian energy. Social Networking, Scientific Research Publishing, v. 2, n. 01, p. 19, 2013. Citado na página 51.

77 SUMMERS, T. H.; LYGEROS, J. Optimal sensor and actuator placement in complex dynamical networks. IFAC Proceedings Volumes, Elsevier, v. 47, n. 3, p. 3784–3789, 2014. Citado na página 52.

78 DALL'ASTA, L. et al. Vulnerability of weighted networks. Journal of Statistical Mechanics: Theory and Experiment, IOP Publishing, v. 2006, n. 04, p. P04006, 2006. Citado 2 vezes nas páginas 52 and 59.

79 ESTRADA, E.; HATANO, N. Statistical-mechanical approach to subgraph centrality in complex networks. Chemical Physics Letters, Elsevier, v. 439, n. 1-3, p. 247–251, 2007. Citado na página 52.

80 DUIJN, P. A.; KASHIRIN, V.; SLOOT, P. M. The relative ineffectiveness of criminal network disruption. Scientific reports, Nature Publishing Group, v. 4, p. 4238, 2014. Citado na página 52.

81 CUNHA, B. R. da et al. Assessing police topological efficiency in a major sting operation on the dark web. Scientific Reports, Nature Publishing Group, v. 10, n. 1, p. 1–10, 2020. Citado na página 52.

82 GRUND, T. U.; DENSLEY, J. A. Ethnic heterogeneity in the activity and structure of a black street gang. European Journal of Criminology, Sage Publications Sage UK: London, England, v. 9, n. 4, p. 388–406, 2012. Citado na página 52.

83 GRUND, T. U.; DENSLEY, J. A. Ethnic homophily and triad closure: Mapping internal gang structure using exponential random graph models. Journal of Contemporary Criminal Justice, SAGE Publications Sage CA: Los Angeles, CA, v. 31, n. 3, p. 354–370, 2015. Citado na página 52.

84 BRANTINGHAM, P. L.; BRANTINGHAM, P. J. Environment, routine and situation: Toward a pattern theory of crime. Advances in criminological theory, v. 5, n. 2, p. 259–294, 1993. Citado 2 vezes nas páginas 52 and 61.

85 CHATURAPRUEK, S. et al. Crime modeling with lévy flights. SIAM Journal on Applied Mathematics, SIAM, v. 73, n. 4, p. 1703–1720, 2013. Citado 2 vezes nas páginas 52 and 61.

86 KREBS, V. Uncloaking terrorist networks. First Monday, 2002. Citado 2 vezes nas páginas 53 and 73.

87 BONACICH, P. Some unique properties of eigenvector centrality. Social networks, Elsevier, v. 29, n. 4, p. 555–564, 2007. Citado na página 53.

88 FREEMAN, L. C. A set of measures of centrality based on betweenness. Sociometry, JSTOR, p. 35–41, 1977. Citado na página 53.

89 SABIDUSSI, G. The centrality index of a graph. Psychometrika, Springer, v. 31, n. 4, p. 581–603, 1966. Citado na página 53.

90 IYER, S. et al. Attack robustness and centrality of complex networks. PloS one, Public Library of Science, v. 8, n. 4, 2013. Citado 3 vezes nas páginas 53, 59, and 67.

91 JAHANPOUR, E.; CHEN, X. Analysis of complex network performance and heuristic node removal strategies. Communications in Nonlinear Science and Numerical Simulation, Elsevier, v. 18, n. 12, p. 3458–3468, 2013. Citado 2 vezes nas páginas 53 and 112.

92 SPARROW, M. K. Network vulnerabilities and strategic intelligence in law enforcement. International Journal of Intelligence and Counter Intelligence, Taylor & Francis, v. 5, n. 3, p. 255–274, 1991. Citado na página 54.

93 CARLEY, K. M. et al. Destabilizing dynamic covert networks. In: Proceedings of the 8th international Command and Control Research and Technology Symposium. [S.l.: s.n.], 2003. p. 79–92. Citado na página 54.

94 SPARROW, M. K. The application of network analysis to criminal intelligence: An assessment of the prospects. Social networks, Elsevier, v. 13, n. 3, p. 251–274, 1991. Citado na página 54.

95 STRANG, S. J. Network analysis in criminal intelligence. In: Networks and network analysis for defence and security. [S.l.]: Springer, 2014. p. 1–26. Citado 2 vezes nas páginas 54 and 60.

96 BORGATTI, S. P. Identifying sets of key players in a social network. Computational & Mathematical Organization Theory, Springer, v. 12, n. 1, p. 21–34, 2006. Citado na página 54.

97 PIRAVEENAN, M.; PROKOPENKO, M.; HOSSAIN, L. Percolation centrality: Quantifying graph-theoretic impact of nodes during percolation in networks. PloS one, Public Library of Science, v. 8, n. 1, 2013. Citado na página 54.

98 MORSELLI, C. Assessing vulnerable and strategic positions in a criminal network. Journal of Contemporary Criminal Justice, SAGE Publications Sage CA: Los Angeles, CA, v. 26, n. 4, p. 382–392, 2010. Citado 2 vezes nas páginas 56 and 78.

99 HAIR, I. T. Principal component analysis and factor analysis. Principal component analysis, Springer, p. 150–166, 2002. Citado na página 61.

100 BRIGHT, D. et al. Criminal network vulnerabilities and adaptations. Global Crime, Taylor & Francis, v. 18, n. 4, p. 424–441, 2017. Citado 2 vezes nas páginas 62 and 67.

101 COHEN, R. et al. Resilience of the internet to random breakdowns. Physical review letters, APS, v. 85, n. 21, p. 4626, 2000. Citado na página 62.

102 COHEN, R. et al. Breakdown of the internet under intentional attack. Physical review letters, APS, v. 86, n. 16, p. 3682, 2001. Citado na página 62.

103 BARABÁSI, A.-L. et al. Network science. [S.l.]: Cambridge university press, 2016. Citado na página 62.

104 MISHKOVSKI, I.; BIEY, M.; KOCAREV, L. Vulnerability of complex networks. Communications in Nonlinear Science and Numerical Simulation, Elsevier, v. 16, n. 1, p. 341–349, 2011. Citado na página 62.

105 KEAN, T.; HAMILTON, L. et al. The 9/11 commission report: Final report of the national commission on terrorist attacks upon the United States. [S.l.]: Government Printing Office, 2004. v. 3. Citado na página 77.

106 BURT, R. S. et al. Brokerage and closure: An introduction to social capital. [S.l.]: Oxford university press, 2005. Citado na página 78.

107 ANDRADE, R. L. d.; RÊGO, L. C. Proposal for the E-I index for non-disjoint groups. The 8th International Conference on Complex Networks and Their Applications. Complex Networks, 2019. Citado 4 vezes nas páginas 80, 92, 93, and 96.

108 ZHANG, J. et al. Data-driven computational social science: A survey. Big Data Research, Elsevier, p. 100145, 2020. Citado na página 80.

109 LEE, D. H.; BRUSILOVSKY, P. How to measure information similarity in online social networks: A case study of citeulike. Information Sciences, Elsevier, v. 418, p. 46–60, 2017. Citado na página 80.

110 BUCUR, D. Gender homophily in online book networks. Information sciences, Elsevier, v. 481, p. 229–243, 2019. Citado na página 80.

111 SAEIDIBONAB, S. Homophily and Friendship Dynamics: An analysis of friendship formation with respect to homophily principle and distinctiveness theory. 2017. Citado na página 80.

112 HIRSCH, J. E. An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. Scientometrics, Springer, v. 85, n. 3, p. 741–754, 2010. Citado 2 vezes nas páginas 85 and 99.

113 BELLA, E. di; GANDULLIA, L.; PRETI, S. Analysis of scientific collaboration network of italian institute of technology. Scientometrics, Springer, v. 126, n. 10, p. 8517–8539, 2021. Citado na página 87.

114 ZADEH, L. A. Fuzzy sets. In: Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh. [S.l.]: World Scientific, 1996. p. 394–432. Citado na página 93.

115 BARNES, M. L. et al. Social networks and environmental outcomes. Proceedings of the National Academy of Sciences, National Acad Sciences, v. 113, n. 23, p. 6466–6471, 2016. Citado na página 109.