sid.inpe.br/mtc-m21d/2022/02.25.14.41-TDI

BEYOND THE SHORTEST PATH: AN ANALYSIS OF NETWORKS' VULNERABILITIES

Giovanni Guarnieri Soares

Master's Dissertation of the Graduate Course in Applied Computing, guided by Dr. Leonardo Bacelar Lima Santos, approved in February 11, 2022.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/46DSFMP>

INPE
São José dos Campos
2022

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

sid.inpe.br/mtc-m21d/2022/02.25.14.41-TDI

BEYOND THE SHORTEST PATH: AN ANALYSIS OF NETWORKS' VULNERABILITIES

Giovanni Guarnieri Soares

Master's Dissertation of the Graduate Course in Applied Computing, guided by Dr. Leonardo Bacelar Lima Santos, approved in February 11, 2022.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/46DSFMP>

INPE
São José dos Campos
2022

Cataloging in Publication Data

Soares, Giovanni Guarnieri.

So11b Beyond the shortest path: an analysis of networks' vulnerabilities / Giovanni Guarnieri Soares. – São José dos Campos : INPE, 2022.

xxii + 84 p. ; (sid.inpe.br/mtc-m21d/2022/02.25.14.41-TDI)

Dissertation (Master in Applied Computing) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2022.

Guiding : Dr. Leonardo Bacelar Lima Santos.

1. Complex networks. 2. Random walks. 3. Vulnerability. 4. Communicability. 5. Eficiency. I.Title.

CDU 004.7

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

DEFESA FINAL DE DISSERTAÇÃO DE GIOVANNI GUARNIERI SOARES
BANCA Nº 026/2022, REG 915763/2020

No dia 11 de fevereiro de 2022, as 14h, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por maioria simples, em cumprimento ao requisito exigido para obtenção do Título de Mestre em Computação Aplicada. O trabalho precisa da incorporação das correções sugeridas pela Banca Examinadora e revisão final pelo(s) orientador(es).

Observação: O aluno foi reprovado pelo Dr. Elbert Einstein Nehrer Macau

Título: "BEYOND THE SHORTEST PATH: AN ANALYSIS OF NETWORKS' VULNERABILITIES"

Membros da banca:

Dr. Marcos Gonçalves Quiles - Presidente - Unifesp
Dr. Leonardo Bacelar Lima Santos - Orientador - Cemaden
Dr. Elbert Einstein Nehrer Macau - Membro Interno - Unifesp
Dr. Thadeu Josino Pereira Penna - - Membro Externo - UFF

Documento assinado eletronicamente por **Marcos Gonçalves Quiles (E)**, **Usuário Externo**, em 14/02/2022, às 19:15 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Leonardo Bacelar Lima Santos**, **Pesquisador**, em 14/02/2022, às 20:25 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Thadeu Josino Pereira Penna (E)**, **Usuário Externo**, em 15/02/2022, às 08:45 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Elbert Einstein nehrer macau (E)**, **Usuário Externo**, em 07/03/2022, às 23:45 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site http://sei.mctic.gov.br/verifica.html, informando o código verificador **9471197** e o código CRC **299DF8E6**.

Referência: Processo nº 01340.001058/2022-04 SEI nº 9471197

"Ford!" he said, "there's an infinite number of monkeys outside who want to talk to us about this script for Hamlet they've worked out." .

Douglas Adams
at *"The Hitchhiker's Guide to the Galaxy"*, 1979

To my parents ***Andrea*** *and* ***Sandro****, my aunts* ***Angela*** *and* ***Fernanda****, and grandmas* ***Isabel*** *and* ***Maria Elizabete****.*

ACKNOWLEDGEMENTS

Gostaria de começar agradecendo à minha mãe, que sempre esteve ao meu lado e me aguentou até nos períodos mais difíceis. Agradecer ao meu pai que sempre me apoiou em tudo que pode e não pode. Agradecer também à minha tia Fernanda, que me orientava mesmo sem saber e é o modelo ideal do cientista no Brasil, o qual tento seguir. Agradecer a minha tia Basa, que sempre que precisei correu pra ajudar e mesmo de longe se fez muito presente. Agradecer aos meus amigos Charlinho, Enila, Elis, Félix, Larissa, Papaco, Sabrinna, Tainara e Thamiris. Alguns que estão comigo há muito tempo, outros que surgiram há pouco, e ainda que voltaram agora. Muito obrigado pelas incontáveis horas de companhia, de D&D, de carinho, força e sanidade mental. Agradeço à minha vó Bete, à minha vó Isabel, e à toda minha família. Agradeço ao meu orientador Leonardo, por me ajudar nessa jornada do mestrado da melhor maneira possível e imaginável. Agradeço aos professores Elbert, Quiles e Thadeu por aceitarem participar da minha banca. Por fim agradeço ao CSILab da UFOP, o qual me disponibilizou um cluster para rodar meus programas com mais facilidade.

I'd like to start thanking my mother, who has always been by my side and supported me even in the most challenging times. Thanks to my father, who always supported me in everything he could and could not. Thanks to my aunt Fernanda, who guides me even without knowing it and is a role model of a scientist in Brazil, which I try to follow. Thanks, Aunt Basa, who helped whenever I needed it and even from far away, was very present. Thanks to my friends Charlinho, Enila, Elis, Félix, Larissa, Papaco, Sabrinna, Tainara and Thamiris. They've been with me for a long time, others for a little while, and some came back later. Thank you so much for the countless hours of companionship, of D&D, of affection, strength, and sanity. I thank my grandmother Bete, my grandmother Isabel, and my family. I would like to thank my advisor Leonardo for helping me on this master's journey in the best possible and imaginable way. I would like to thank professors Elbert, Quiles, and Thadeu for agreeing to review my dissertation. Finally, I thank UFOP's CSILab, which provided a cluster to run my programs quickly and precisely.

This research was partially supported by grant 420338/2018-7 of the Brazilian National Council for Scientific and Technological Development (CNPq) and by grants 2018/06205-7 and 420338/2018-7 of São Paulo Research Foundation (FAPESP) and DFG-IRTG 1740/2

ABSTRACT

Transportation, power grids, communication, water, oil, and gas distribution systems heavily influence our well-being. Studying the structure of those critical infrastructures is extremely important for people's quality of life. We model those systems as a network using the graph theory. This work explores already defined topological metrics such as Efficiency, Communicability, and Vulnerability, but following an innovative approach: considering not only the shortest paths between pairs of nodes in the networks. We define vulnerability as a drop in the network's performance, and performance is a general term, able to be quantified by different metrics. We propose a twist in the already defined vulnerability index using communicability as a performance instead of efficiency. Firstly, we compare the traditional efficiency-based vulnerability with our proposed communicability-based one. This way, we show how the different metrics highlight different vulnerable points and how testing multiple paths instead of only the shortest can impact the results. After that, we perform several linear regressions between the vulnerabilities and well-known metrics (e.g., degree, shortest path length, and betweenness). Our findings show different patterns of relations for different network topologies, such as Random and Scale Free. Finally, we explore Random Walks on networks by walking on them. We run a Random Walk on the network and count how many times the Brownian particle goes through each node: Passaging Index. We compare, for each node, the Passaging Index with other metrics and find linear correlations between the Passaging Index and Degree, resulting in a strong linear correlation, with a $R^2 = 1$. To the Erdős–Rényi model, the linear correlation is also present between the Passaging Index and the Vulnerability with Efficiency in all cases, while in Barabási-Albert presents more complex correlations between metrics.

Keywords: Complex Networks. Random Walks. Vulnerability. Communicability. Eficiency.

ALÉM DO CAMINHO MAIS CURTO: UMA ANÁLISE DAS VULNERABILIDADES DAS REDES

RESUMO

Transporte, redes de energia, comunicação e sistemas de distribuição de água, óleo e gás influenciam fortemente nosso bem-estar. Estudar a estrutura dessas infraestruturas críticas é extremamente importante para a qualidade de vida das pessoas. Modelamos esses sistemas como uma rede usando a teoria dos grafos. Este trabalho explora métricas topológicas já definidas como Eficiência, Comunicabilidade e Vulnerabilidade, mas seguindo uma abordagem inovadora: considerando não apenas os caminhos mais curtos entre pares de nós nas redes. Definimos vulnerabilidade como uma queda na performance da rede, e performance é um termo geral, passível de ser quantificado por diferentes métricas. Propomos uma reviravolta no índice de vulnerabilidade já definido usando comunicabilidade como um desempenho em vez de eficiência. Em primeiro lugar, comparamos a vulnerabilidade tradicional baseada na eficiência com a nossa proposta baseada na comunicabilidade. Desta forma, mostramos como as diferentes métricas destacam diferentes pontos vulneráveis e como testar vários caminhos em vez de apenas o mais curto pode impactar os resultados. Depois disso, realizamos várias regressões lineares entre as vulnerabilidades e métricas conhecidas (por exemplo, grau, comprimento do caminho mais curto e intermediação). Nossos resultados mostram diferentes padrões de relações para diferentes topologias de rede, como Aleatória e Livre de Escala. Finalmente, exploramos Random Walks em redes caminhando sobre elas. Executamos uma Random Walk na rede e contamos quantas vezes a partícula browniana passa por cada nó: Índice de Passagem. Comparamos, para cada nó, tal índice com outras métricas de rede e encontramos uma forte correlação linear entre o Índice de Passagem e o grau associado a cada nó, formando uma relação direta, com $R^2 = 1$ para todos os casos de grafos distintos. Dentre estes cálculos, também foi encontradas outras relações lineares, entre o Índice de Passagem e a Vulnerabilidade com Eficicência, na relação com o modelo de Erdős–Rényi. Esta linearidade não se mantém tão forte no modelo de Barabási-Albert, apresentando outras relações mais complexas.

Palavras-chave: Redes Complexas. Random Walks. Vulnerabilidade. Comunicabilidade. Eficiência.

LIST OF FIGURES

Page

2.1 Scheme of the Königsberg bridges. 5
2.2 Graphic representation of the Königsberg bridges. 6
2.3 Example graph to common metrics. 8
2.4 Graphical representation of Erdős–Rényi model used in this work. It's generated with 100 nodes, 1475 edges, and a probability of wiring $p = 0.3$. 10
2.5 Graphical representation of Barabási-Albert model used in this work. It was generated with 100 nodes, 1547 edges, and a number of outgoing edges $m = 17$. 11

3.1 Flowchart illustrating how to calculate the vulnerability to a given graph. In this work, we will use global efficiency and communicability as performance. 18
3.2 Zachary's Karate Club graph, used to test our results. 19
3.3 Visualizing the point-wise vulnerability of the Zachary Karate Club. This vulnerability is calculated using efficiency as performance. 20
3.4 A visualization of the point-wise vulnerability of the Zachary Karate Club. This vulnerability is calculated using communicability as performance. 21

4.1 Scatter plots to graph with following characteristics: N = 100, L = 1475 d = 0.30, $< c >$ = 0.30, $< k >$ = 29.50, D = 2, $< l >$ = 1.68. 28
4.2 Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Figures 2.4, 2.5. 29
4.3 Erdős–Rényi distribution of each vulnerability shows how many times a value inside the bin's interval appears. The line represents the sample's mean value. 30
4.4 Scatter plots to graph with following characteristics: N = 100, L = 584, d = 0.11, $< c >$ = 0.11 , $< k >$ = 11.68, D = 4, $< l >$ = 2.08. 31
4.5 Scatter plots to graph with following characteristics: N = 256, L = 1329, d = 0.04, $< c >$ = 0.04, $< k >$ = 10.38, D = 4, $< l >$ = 2.61. 32
4.6 Scatter plots to graph with following characteristics: N = 2000, L = 99685, d = 0.05, $< c >$ = 0.05, $< k >$ = 99.68, D = 3, $< l >$ = 1.95. . . 33
4.7 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$. 34

4.8 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$. 35
4.9 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$. 36
4.10 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$. 37
4.11 Scatter plots to graph with following characteristics: N = 100, L = 1547, d = 0.3, $< c > = 0.42$, $< k > = 30.94$, D = 3, $< l > = 1.67$. 38
4.12 Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Figures 2.5, 2.5. 39
4.13 Barabási-Albert distribution of each vulnerability shows how many times a value inside the bin's interval appears. The line represents the sample mean value. 40
4.14 Scatter plots to graph with following characteristics: N = 100, L = 672, d = 0.13, $< c > = 0.23$, $< k > = 13.44$, D = 3, $< l > = 1.99$. 41
4.15 Scatter plots to graph with following characteristics: N = 256, L = 1515, d = 0.05, $< c > = 0.10$, $< k > = 11.83$, D = 4, $< l > = 2.45$. 42
4.16 Scatter plots to graph with following characteristics: N = 2000, L = 98725, d = 0.05, $< c > = 0.12$, $< k > = 98.72$, D = 3, $< l > = 1.97$. . . 43
4.17 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$. 44
4.18 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$. 45
4.19 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$. 46
4.20 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$. 47
4.21 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$. 48
4.22 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$. 49
4.23 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$. 50
4.24 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$. 50
4.25 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$. 51

4.26 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$. 52
4.27 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$. 52
4.28 Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$. 53

5.1 Flowchart representing how the code works. 56
5.2 Color map illustrating the Passaging Index to each Zachary Karate Club network node. 58
5.3 Standard Deviation calculated to compare the arrays made with the Passaging Index of each node. Where μ represents 10^{-6}. Made with the Zachary's Karate Club. 59
5.4 Zachary's Karate Club Passaging Index relations. 59
5.5 Scatter plots to graph with following characteristics: N = 100, L = 1475 Density = 0.3, N = 100, L = 1547, $< c > = 0.3$, $< k > = 29.50$, D = 2, $< l > = 1.68$. 60
5.6 Scatter plots to graph with following characteristics: N = 100, L = 1547 Density = 0.3, N = 100, L = 1547, $< c > = 0.42$, $< k > = 30.94$, D = 3, $< l > = 1.67$. 61

A.1 Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4. 69
A.2 Erdős–Rényi distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value. 70
A.3 Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4. 71
A.4 Erdős–Rényi distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value. 72
A.5 Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4. 73
A.6 Erdős–Rényi distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value. 74

A.7 Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4. 75
A.8 Barabási-Albert distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value. 76
A.9 Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4. 77
A.10 Barabási-Albert distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value. 78
A.11 Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4. 79
A.12 Barabási-Albert distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value. 80

B.1 Scatter plots to graph with following characteristics: N = 100, L = 584, Density = 0.11, $< c > = 0.11$, $< k > = 11.68$, D = 4, $< l > = 2.08$. . . 81
B.2 Scatter plots to graph with following characteristics: N = 256, L = 1329, Density = 0.04, $< c > = 0.04$, $< k > = 10.38$, D = 4, $< l > = 2.61$. . . . 82
B.3 Scatter plots to graph with following characteristics: N = 2000, L = 99685, Density = 0.05, $< c > = 0.05$, $< k > = 99.68$, D = 3, $< l > = 1.95$. 82
B.4 Scatter plots to graph with following characteristics: N = 100, L = 672, Density = 0.13, $< c > = 0.23$, $< k > = 13.44$, D = 3, $< l > = 1.99$. . . . 83
B.5 Scatter plots to graph with following characteristics: N = 256, L = 1515, Density = 0.05, $< c > = 0.10$, $< k > = 11.83$, D = 4, $< l > = 2.45$. . . . 84
B.6 Scatter plots to graph with following characteristics: N = 2000, L = 98725, Density = 0.05, $< c > = 0.12$, $< k > = 98.72$, D = 3, $< l > = 1.97$. 84

LIST OF ABBREVIATIONS

N	–	Nodes;
L	–	Edges or links
RW	–	Random Walk;
WWW	–	World Wide Web;
Vul(Efi)	–	Vulnerability with Efficiency as performance;
Vul(Com)	–	Vulnerability with Communicability as performance;
k_i	–	Node's Degree;
d	–	Graph's density;
g_k	–	Nodes's Betweenness;
$< l >_i$	–	Node's Mean Shortest Path;
PI	–	Passaging Index.

CONTENTS

Page

1 INTRODUCTION 1
1.1 Text organization 3

2 THEORETICAL BACKGROUND 5
2.1 Graphs 5
2.2 Graph models 9
2.3 Walks and paths 12
2.4 Literature overview 12
2.4.1 Vulnerability and metrics 13
2.4.2 Random walks 14

3 STUDYING THE PERFORMANCE DROP USING COMMUNICABILITY 15
3.1 Introduction 15
3.2 Metrics 16
3.2.1 Efficiency 16
3.2.2 Communicability 16
3.2.3 Vulnerability 17
3.3 Finding vulnerabilities 18
3.4 Results 20
3.5 Conclusion 22

4 VULNERABILITIES AND HOW THEY RELATE WITH OTHER METRICS 25
4.1 Introduction 25
4.2 Methodology 25
4.3 Results 27
4.3.1 Erdős–Rényi model 27
4.3.1.1 Scatter plots 27
4.3.1.2 Density variation 33
4.3.2 Barabási-Albert model 37
4.3.2.1 Scatter plots 37
4.3.2.2 Density variation 43

4.4 Conclusion 53

5 RANDOM WALKS: A STOCHASTIC VIEW ON THE HIERARCHY OF NODES 55

5.1 Introduction 55
5.2 Methodology 55
5.3 Results 57
5.4 Conclusion 62

6 FINAL REMARKS 63

REFERENCES 65

APPENDIX A - ADDITIONAL FIGURES FROM CHAPTER 4 . . 69

A.1 Erdős–Rényi 69
A.1.1 N = 100, L = 584 69
A.1.2 N = 256, L = 1329 71
A.1.3 N = 2000, L = 99685 73
A.2 Barabási-Albert 75
A.2.1 N = 100, L = 584 75
A.2.2 N = 256, L = 1329 77
A.2.3 N = 2000, L = 99685 79

APPENDIX B - ADDITIONAL FIGURES FROM CHAPTER 5 . . 81

B.1 Erdős–Rényi 81
B.2 Barabási-Albert 83

1 INTRODUCTION

The study of Complex Networks is increasingly present in the sphere of researchers worldwide. The theory of networks arose in a study by Leonahrd Euler in the mid-1730s (EULER, 1736) (original in Latin), and it took until 1959 to begin to grow and spread to other sciences (ERDŐS; RÉNYI, 1959; ERDőS; RéNYI, 1960; WATTS; STROGATZ, 1998; ALBERT; BARABÁSI, 2002; BARABÁSI; BONABEAU, 2003). Networks are a typical structure in our lives; they permeate nature and society. The people you know are part of a social network. Your computer connected to the internet is part of a worldwide computer network. A food chain is simply a network of who feeds on whom food relationships.

The complex networks approach is currently present in several areas, such as Neuralscience (HOPFIELD, 1982), biology (JEONG et al., 2000), climate (TSONIS et al., 2006) and even in natural disaster's impacts (SANTOS et al., 2015).

The United Nations Office for Disaster Risk Reduction defines disasters as "A serious disruption of the functioning of a community or a society at any scale due to hazardous events interacting with conditions of exposure, vulnerability and capacity, leading to one or more of the following: human, material, economic and environmental losses and impacts.", and vulnerability as "The conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards." (UNDRR, 2022).

An example case of a disaster exposing a vulnerability happened in 2020. A blackout that generated an energy crisis in the state of Amapá, where almost 90% of the state was left without power after a fire reached the main power substation in the state (G1, 2020). Also in 2021 the failure was caused by excess demand on a freezing winter day in Texas, US. The freezing day froze the natural gas and blocked the pipes that carried them, froze the wind turbines and coal piles, making it impossible for supply to catch up with demand. This accident happened because the state of Texas is not connected to the rest of the country's power grid (GUARDIAN, 2021). This way, we need to anticipate these problems and reduce the disruption caused by disasters.

Analyzing the entire structure as a network, locating critical points, and discussing the causes of greater vulnerability for the network as a whole is where Complex Networks come in this work. There are already some ways to do that kind of analysis of finding the most vulnerable structures in a network by looking into its topology

in a few different ways (HOLME et al., 2002; MISHKOVSKI et al., 2011; CHEN et al., 2010; ROCCO; RAMIREZ-MARQUEZ, 2011). In this work, we both show one of the types (GOLDSHTEIN et al., 2004) and define a variation since the original definition is flexible to (LATORA; MARCHIORI, 2004) changes.

Almost all of those vulnerability metrics have something in common: they use the shortest path in their calculation. A path is simply a route or track between one place and another, and the shortest path is the path with shorter length in the set of possible paths. The shortest path is often seem as the the most efficient one, since efficiency is closely related to the least amount of time spent when executing a task.

Taking the shortest (or most efficient) path is what we assume to be the most common practice, but it does not always happen; people do not always follow the lowest cost path, as shown by the work of (LIMA et al., 2008) using the data from 92419 GPS trajectories. Besides, the shortest path is not always the safest path, as shown in (GALBRUN et al., 2016), where they develop an algorithm to output a small set of paths providing tradeoffs between distance and safety.

This brings to light our scientific question: how does considering not only the shortest paths can affect the vulnerability index in complex networks?

As a tool, we have the communicability (ESTRADA; HATANO, 2008) which, in contrast to efficiency, presents analyses seeking to evaluate all the paths that lead from one point to another, rather than just the shortest path.

Communicability is an analytical solution that is very similar to a diffusive process in the network as if we transformed the network into a set of springs and when moving one of them, we verified how all the others start to move. On another side of the same coin, we have Random Walks, stochastic processes resembling Brownian motion, which arises when discussing diffusion in physics. Both communicability and Random Walks have this connection with the diffusive process, and when simulating Random Walks the path is going to repeat some nodes, just as communicability considers all possible paths, even repeating points.

This work aims to study metrics already defined and common in complex networks, such as efficiency, communicability, and vulnerability. It is interesting to note that vulnerability is defined in a very general way (LATORA; MARCHIORI, 2004) and is implemented in (GOLDSHTEIN et al., 2004) using efficiency as performance. By inspiration, our prime objective is defining vulnerability using communicability as

performance. Lastly, by using Random Walks, we incorporate the non-shortest paths with a stochastic point of view.

1.1 Text organization

We organize our text in Chapters, being the first one this introduction and the second a brief theoretical background. Chapters 3, 4, and 5 are related works, where we begin by proposing a metric, testing its characteristics, then trying to relate it with a new stochastic version. In the last Chapter, we talk about our conclusions and future prospects.

- Chapter 2: Basic concepts of graph, graph models, and walks & paths. Definition and showcases of each subject, with a short history background of graphs. We close the chapter going through a literature overview, treating the main points inside our work.
- Chapter 3: This Chapter is an already published work on Vulnerability with Communicability. This work brings a twist in an already existing vulnerability metric but changes the performance from efficiency to communicability, defining the Vulnerability with Communicability metric.
- Chapter 4: This Chapter aims further explore our previous work by comparing its result to two different models of graphs with already well-established metrics in literature. The Chapter shows how each metric relates with both vulnerabilities, to both models and 4 different graphs.
- Chapter 5: Here we explore a different type of metric by incorporating the non-shortest path as a stochastic index, here called as Passaging Index.
- Chapter 6 Lastly, we conclude our work with our final remarks, where we explain our conclusions and what to expect from future works.

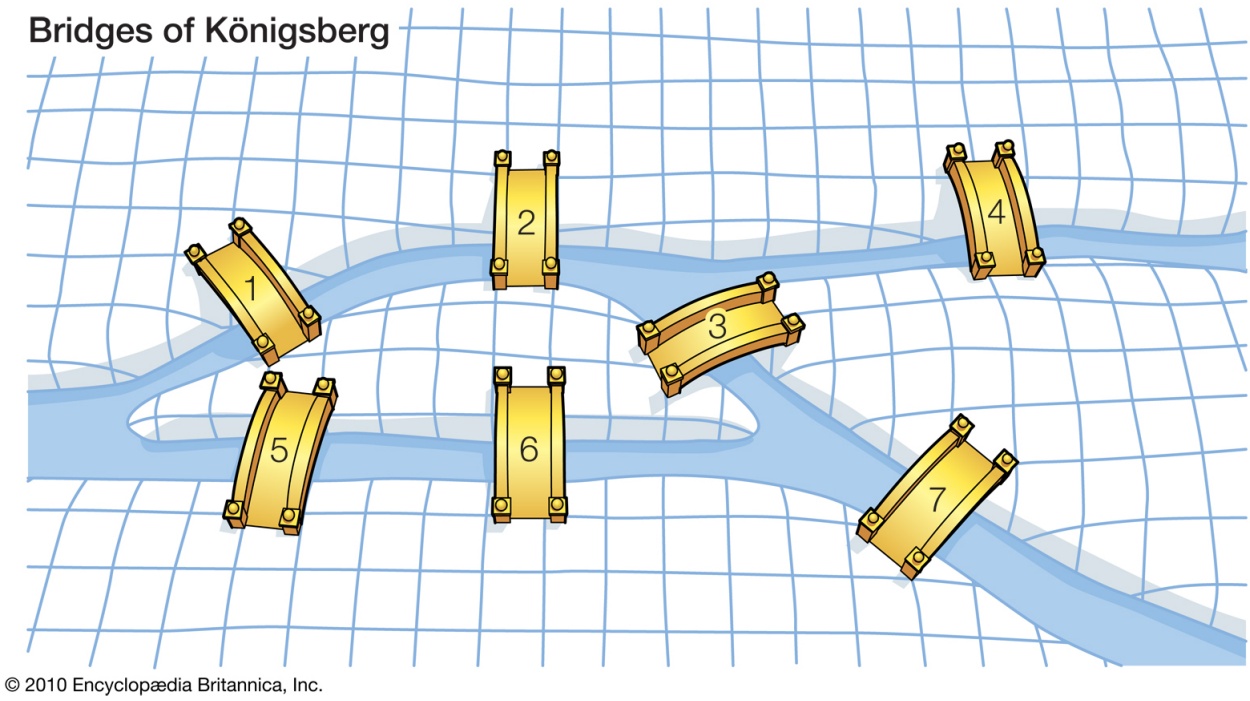
2 THEORETICAL BACKGROUND

This Chapter explains the essential theoretical background needed to comprehend our work. We define the mathematical tool of graphs with history and formal definitions; then, we start an overview of the two most traditional models of graphs, finishing with an explanation of walks and paths.

2.1 Graphs

Euler, a Swiss mathematician who spent most of his career in Berlin and St. Petersburg, had an extraordinary influence in all areas of mathematics, physics, and engineering, not just in quantity but also in quality. In the town of Königsberg, close to St. Petersburg, arose a simple question: can anyone walk across all seven bridges and never cross a bridge again?

Figure 2.1 - Scheme of the Königsberg bridges.



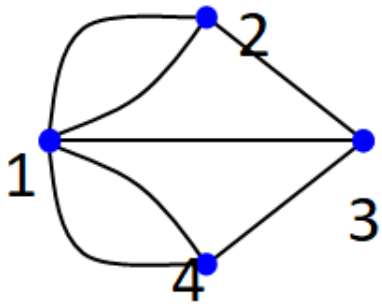
Source: Carlson (2021).

Euler said that there was no possible way to do this, and subtly, he started a new branch of mathematics known today as graph theory. Graph theory today is the basis for our study of networks, and this theory only began to expand into something

more concrete centuries after Euler, when great mathematicians began to study the phenomenon and helped open doors in the new field of complex networks (BARABáSI, 2002).

Euler had a simple, elegant thought and was easily understood even by those not trained in mathematics. The study was not a complex development, but the tool he created to develop this solution made this turnaround. Euler abstracted from the idea of physical space and made the following representation:

Figure 2.2 - Graphic representation of the Königsberg bridges.



Source: WIKIPEDIA (2021).

This set of dots with connections between them is what we call a graph, the dots are nodes (or vertices), and the connections are edges. Mathematically speaking, we have

$$G = (N, L), \tag{2.1}$$

where G is the set of the graph, N is the set of nodes, and L is the set of edges (or links).

This simple idea helped to notice that there is no way to cross the seven bridges just once. To cross it only once, the nodes with an odd number of edges should be either the beginning or the end of the walk, with a limit of two or no node with an odd number of edges. As in Figure 2.2, all four nodes have an odd number of bridges, so it is not possible to make the path (BARABáSI, 2002).

Later was developed the graph structure to a more mathematical view; with the

representation of a matrix, we can map the connections of each node, as follows

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}, \tag{2.2}$$

this is the adjacency matrix of Figure 2.2 following a clockwise order beginning from the left, where the element A_{ij} of the matrix A is equal to 1 if there is a connection (edge) and equal to 0 when there is not.

Using this definition, we can extend to find some graph, nodes, and edges properties:

- Degree (k_i): which is the number of edges a node has, connecting it to other nodes. Can be calculated using the adjacency matrix with the following equation

$$k_i = \sum_{j=1}^{N} A_{ij}, \tag{2.3}$$

  where k_i is the node's degree, A_{ij} is an element of the adjacency matrix. In short, sum all elements of the i_{th} line.

- Density (d): Density is a measure of how many connections a network has divided by how many it could have.

$$d = \frac{Numberofedges}{Possiblenumberofedges}. \tag{2.4}$$

  Where,

$$Possiblenumberofedges = \frac{N \times (N-1)}{2} \tag{2.5}$$

- Shortest path length (l_{ij}): The length of the path with the minimum number of edges connecting nodes i and j. It can also be called distance between nodes i and j.

- Node's mean shortest path ($< l >_i$): The mean of shortest paths leaving node i to all others. It is defined as

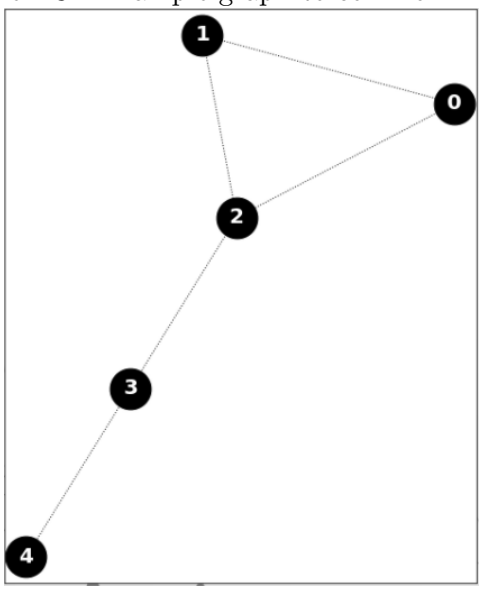
$$< l >_i = \frac{1}{N-1} \sum_{i=1, j \neq i}^{N} l_{ij}. \tag{2.6}$$

- Betwenness (g_k): The node betweenness is how many shortest paths go through the node i. It is defined as

$$g_k = \sum_{i \neq j \neq k} \frac{\sigma_{ij}(k)}{\sigma_{ij}}, \tag{2.7}$$

where σ_{ij} is the number of shortest paths going from i to j and $\sigma_{ij}(k)$ is the number of paths that goes through k.

Figure 2.3 - Example graph to common metrics.



Using Figure 2.3 as an example, we build the following adjacency matrix.

$$B = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{2.8}$$

With its adjacency matrix in Equation 2.8, we build a table with each metric explained before.

Table 2.1 - Table showing the calculated metrics to each node from the graph represented in Figure 2.3. k_i is the node's degree, $< l >_i$ stands for the node's Mean Shortest Path and g_i is the node's Betweenness.

Node	k_i	$< l >_i$	g_i
0	2	1.75	0
1	2	1.75	0
2	3	1.25	8
3	2	1.5	6
4	1	2.25	0

This data explains that the node 2 has more connections than others and can easily access them, also has the greatest number of shortest paths going through.

2.2 Graph models

There are two very important topologies of networks: random and scale-free networks. To simulate those ones, we randomly generate graphs - in this work, we use the Erdős–Rényi and Barabási-Albert models representing the Random Network and Scale-free network, respectively.

The first introduced model to randomly generate a network was the Erdős–Rényi model in 1959 (ERDŐS; RÉNYI, 1959; ERDőS; RéNYI, 1960). Paul Erdős and Alfréd Rényi proposed a model where given N links, how many of them were linked together with a probability of p.

If we have a network with ten nodes, we choose node 0 as a start and toss a weighted coin to form a connection between it and the other nine nodes, then to node 1 and all the other eight nodes until there are no more nodes work through. Setting the probability to 100%, we will end with a complete graph; setting it to 0%, we will end with a graph with no connections.

Then came Albert-László Barabási and Réka Albert with their study of the World Wide Web, expecting to find a random network. To their surprise, they found out that the WWW was held together by a few highly connected pages. More than 80 percent of the pages on the map had fewer than four links, but a small minority, less than 0.01 percent, had more than a thousand (BARABÁSI; ALBERT, 1999).

So, they came up with a Scale-free network model, inspired by the idea of growth and preferential attachment. The Erdős–Rényi model starts with N nodes and ends

with N nodes, rewiring or not the edges between those nodes. In the Barabási-Albert model, the network starts with a small number of nodes and, with every time-step, adds a new node with a set amount of edges. Those edges have a preferential attachment when choosing the nodes to which the new node connects, proportional to the other nodes' degree and inverse to the total degree of the network (ALBERT; BARABÁSI, 2002).

Figure 2.4 - Graphical representation of Erdős–Rényi model used in this work. It's generated with 100 nodes, 1475 edges, and a probability of wiring $p = 0.3$.

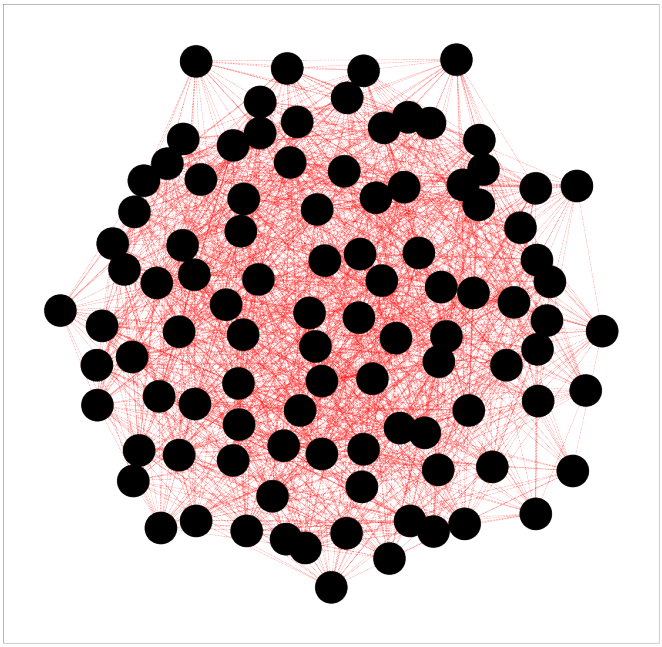
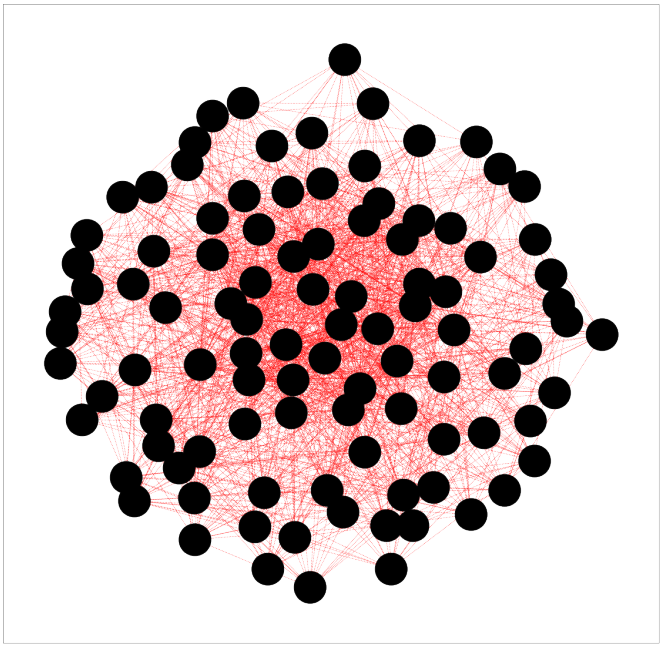


Figure 2.5 - Graphical representation of Barabási-Albert model used in this work. It was generated with 100 nodes, 1547 edges, and a number of outgoing edges $m = 17$.



In Figures 2.4 and 2.5, we can see the topographical difference between each model and it is essential to notice that both graphs have a similar amount of nodes and edges.

Each model has its signature, and this difference becomes apparent when comparing their metrics. When investigating the degree of each node in those models, we can see different distributions taking place. The Erdős–Rényi model, simulating a random network, has a Poisson (or binomial) degree distribution, while the Barabási-Albert model simulates a scale-free network presenting a power-law degree distribution.

Since most nodes in a scale-free network have a low degree, random failures do not impact the network significantly. Contrarily, target attacks may produce huge impacts, reaching few but highly connected nodes. While in random networks, there are few nodes with few connections. When comparing one to another, we conclude that random networks are more susceptible to failures while scale-free is more susceptible to attacks.

2.3 Walks and paths

A set of vertex $v_0, v_1, v_2, ..., v_n$ is called a path if there are no repeated vertexes or walk if there are. This way, a path is a particular case of a walk (BENDER, 2015).

The power of the adjacency matrix A^k mentioned above is a crucial tool when considering walks; it calculates the number of walks starting in node i, ending in node j with k steps.

$$A^1 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}, \tag{2.9}$$

$$A^2 = \begin{bmatrix} 3 & 1 & 2 & 1 \\ 1 & 2 & 1 & 2 \\ 2 & 1 & 3 & 1 \\ 1 & 2 & 1 & 2 \end{bmatrix}, \tag{2.10}$$

$$A^3 = \begin{bmatrix} 4 & 5 & 5 & 5 \\ 5 & 2 & 5 & 2 \\ 5 & 5 & 4 & 5 \\ 5 & 2 & 5 & 2 \end{bmatrix}. \tag{2.11}$$

Then, the element A^2_{00} from Equation 2.10 means that if you start a walk at node 0, there are 3 distinct ways of returning to the node 0 taking 2 steps, if we use the Equation 2.11, there are 4 distinct ways of returning taking 3 steps. However, the number of paths between two nodes is not analytically defined as the number of walks is.

We develop the idea of walks in the metric Communicability, modeling a diffusion-like process inside the network; meanwhile, the concept of paths is not simple to extend in other metrics since it is not easy to represent them mathematically.

2.4 Literature overview

This section gives an overview of the literature used in this work, explaining how each of the main articles helped build this dissertation. There are two main topics,

Vulnerability and metrics, in which we explain where the main metrics came from in chronological order, and Random walks, where we talk about random walks in networks, exploring networks by walking on them.

2.4.1 Vulnerability and metrics

Starting with the main metrics used in this work, we go through Efficiency. Efficiency, as we use it, is first defined in (LATORA; MARCHIORI, 2001) as a measure of how efficiently it exchanges information. In (LATORA; MARCHIORI, 2001), they apply the idea of Efficiency to neural networks and manufactured communication and transportation systems. They show how a network with small-world characteristics is highly efficient since they are globally and locally efficient.

Later, in (LATORA; MARCHIORI, 2004), arises the definition of topological Vulnerability of critical infrastructures. This definition is a general method to spot the critical components of a network. Their method is not only to find vulnerabilities and can help to improve and better shape an expansion in networks. It is an approach that damages or improves the network, checking a drop or increase in performance.

Performance can be a plenitude of different metrics, and in that work, they use Efficiency as performance. The definition of Vulnerability we use in this work was later given by (GOLDSHTEIN et al., 2004), where they use Efficiency as the inverse of the shortest path length from i to j as performance and global Efficiency as a mean of this value. They bring to light the point-wise Vulnerability, which is the Vulnerability associated with each node inside the network. It measures how the removal of the said node can make a direct drop in the network's Efficiency. They define Vulnerability in a proposal to study the hierarchy of networks.

Later on, (ESTRADA; HATANO, 2008) defines Communicability as a new measure of complex networks. Communicability is a broad generalization of the concept of the shortest path. They use this information to distinguish finer structures of networks, such as the communities dividing the network. Communicability is interesting since it takes non-shortest walks into account, with appropriate weights, where long walks have lower contributions to the Communicability function.

In (MISHKOVSKI et al., 2011), they consider the normalized average edge betweenness of a network as a type of Vulnerability.

Later, in (CHEN et al., 2021), they develop a new method for how to model and assess Vulnerability in transit networks. This idea is a hybrid approach integrating

disaster chains and complex networks while using Efficiency as an anchor for further risk improvement.

2.4.2 Random walks

To our random walks chapter, we used as an inspiration two main works. Both of them uses the idea of exploring the graph via walking on it, releasing a walker inside and analyzing its trajectory to certain objectives.

In (YANG, 2005) they compare searches strategies adopted by the walker, using Random Walks with no memory, No-Back Walk, No-Triangle-Loop Walk, No-Quadrangle-Loop Walk, and Self-Avoiding Walk. The idea is increasing the walker's memory to test the best strategy in leaving a node i to a node j. In the end, they find that dynamical processes on networks are greatly dependent on the topological features of the networks.

Further, in (COSTA; TRAVIESO, 2007) they explore the network with different types of random walks with the intent to infer topological measures from these types of walks. They find from a sample of the network values as the average node degree and average cluster coefficient for the entire network.

3 STUDYING THE PERFORMANCE DROP USING COMMUNICABILITY

In this Chapter, we will propose a new way to calculate the Vulnerability of a network and then compare it to an already common one. We define Vulnerability by a drop in the network's performance, and performance is a broad term meaning other metrics can fit in. First, we explore Vulnerability with efficiency as performance; then, we define Vulnerability with communicability as performance. Each one shows different results when studying the Zachary Karate Club network.

The introduction is in Section 3.1 leading to a brief explanation on the vulnerability, efficiency, and communicability metrics in 3.2. Then, we explain how to find the vulnerabilities in 3.3 and show the results in 3.4 to conclude in 3.5.

This work is based on a complete paper published at Encontro Nacional de Modelagem Computacional (ENMC) 2021 (SOARES; SANTOS, 2021).

3.1 Introduction

Complex Networks permeates through science, being applied to a vast range of subjects due to the presence of Networks in nature. Those subjects can be Neural Networks (HOPFIELD, 1982), metabolic networks (JEONG et al., 2000), climate networks (TSONIS et al., 2006), and infrastructure networks (SANTOS et al., 2015).

One of complex network's applications is finding critical points, and identifying the vulnerable nodes to the network as a whole. There are already some ways to find vulnerabilities by studying the topology (HOLME et al., 2002; MISHKOVSKI et al., 2011; CHEN et al., 2010; ROCCO; RAMIREZ-MARQUEZ, 2011) and there are some work with real world data (SANTOS et al., 2020). In this work, we show one of them (GOLDSHTEIN et al., 2004) and succeed in defining a new one, since the original definition is broad and can work very well with many metrics (LATORA; MARCHIORI, 2004).

The common factor between those vulnerabilities metrics is the use of the shortest path. Taking the shortest path, or the most efficient one, is the systematic manner to transit but is not the only way to do so, as shown in (LIMA et al., 2008). Studying 92419 GPS trajectories describing the movement of personal cars over 18 months, they found that a significant fraction of driver's routes is not optimal. Those individual routing choices are not captured by path optimization, but their spatial bounds are similar, even for trips performed by distinct individuals and at various

scales (LIMA et al., 2008).

Communicability is a broad generalization of the concept of the shortest path. It has the characteristic of using all possible ways between two nodes instead of the shortest one. That different approach can arise desirable features depending on the network being treated (ESTRADA; HATANO, 2008).

In this work, we explore the differences when calculating a network vulnerability. We demonstrate how considering multiple paths instead of only the shortest one can result in very different analyses and shed some light on many other spots of vulnerability using the well-known Zachary's Karate Club network.

3.2 Metrics

Here we define the metrics used in this work, giving the analytical definitions and where they came from with a short explanation.

3.2.1 Efficiency

Here we use efficiency as the inverse of the shortest distance length (distance) between two nodes, so the efficiency between the nodes i and j is defined as follows

$$E_{ij} = \frac{1}{d_{ij}}, \tag{3.1}$$

where d_{ij} is the size of the shortest path between both nodes i and j.

Global efficiency is also defined as

$$E(\mathbf{G}) = \frac{1}{N(N-1)} \sum_{i \neq j \in \mathbf{G}} \frac{1}{d_{ij}}; \tag{3.2}$$

where $E(G)$ is the global efficiency of a network G, N is the number of nodes, and d_{ij} is the shortest path between two nodes i and j (LATORA; MARCHIORI, 2001).

3.2.2 Communicability

Communicability is one of the metrics in Complex Networks that aims to model the propagation through the network. This type of propagation does not always take the shortest path but walks inside the network between nodes. We calculate

Communicability using the sum of the number of walks between two nodes

$$Com_{ij} = \sum_{k=0}^{\infty} c_k \#walks. \tag{3.3}$$

The #walks (number of walks) is related to the power of the network's adjacency matrix. The adjacency matrix A gives us the connections between nodes, if the element A_{ij} is 0, there is no edge between the nodes i and j, and if the element A_{ij} is 1, there is a connection between them.

The element c_i needs to make the series convergent and give more weight to shorter paths, so we turn it into a factorial, and the following Equation for Communicability between two nodes is defined

$$Com_{ij} = \sum_{k=0}^{\infty} \frac{A_{ij}^k}{k!} = e_{ij}^{\mathbf{A}}, \tag{3.4}$$

where A_{ij}^k is the element ij of the power of the adjacency matrix, and Com_{ij} is the communicability between both nodes i and j. The power of the adjacency matrix gives us the total number of ways between two nodes. They define communicability between them by giving more relevance to shortest walks and less to longest ones. The shortest path connecting two nodes always gives us the most significant contribution to communicability, but longer walks (which can be more abundant) also have their contribution (ESTRADA; HATANO, 2008).

3.2.3 Vulnerability

Vulnerability is broadly defined, where we can input other very different metrics as performance, giving us different results. It is defined as follows

$$V[S, D] = \frac{\Phi[S] - W[S, D]}{\Phi[S]}, \tag{3.5}$$

where the vulnerability V when a structure S is damaged with D changing its performance Φ and W[S, D] = $\Phi[damage(S, d)]$ is the worst performance of S under the class of damages D (LATORA; MARCHIORI, 2004).

The point-wise vulnerability is defined as

$$V(i) = \frac{E(G) - E(G,i)}{E(G)}, \tag{3.6}$$

where $V(i)$ is the point-wise vulnerability of the node i, $E(G)$ is the global efficiency of the network, as defined in (LATORA; MARCHIORI, 2001), and $E(G,i)$ is the global efficiency of a new similar graph, being the only difference is the node i disconnection from the network (GOLDSHTEIN et al., 2004).

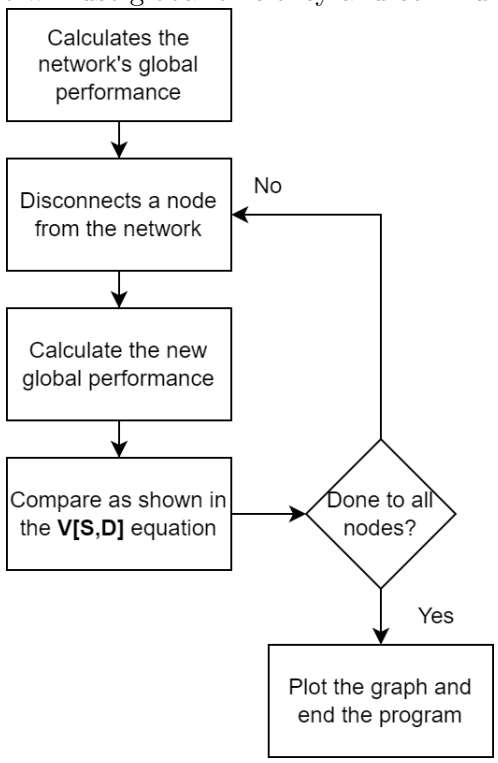
Vulnerability is calculated to the entire network, but is related to the disconnection of each element of the network.

3.3 Finding vulnerabilities

To calculate the network vulnerability, as defined before, we need to go through every node disconnecting it from the network one by one, verifying how this node affects the global performance by recalculating it and comparing the original one with the new one, as seen in Figure 3.5.

The algorithm is as follows

Figure 3.1 - Flowchart illustrating how to calculate the vulnerability to a given graph. In this work, we will use global efficiency and communicability as performance.



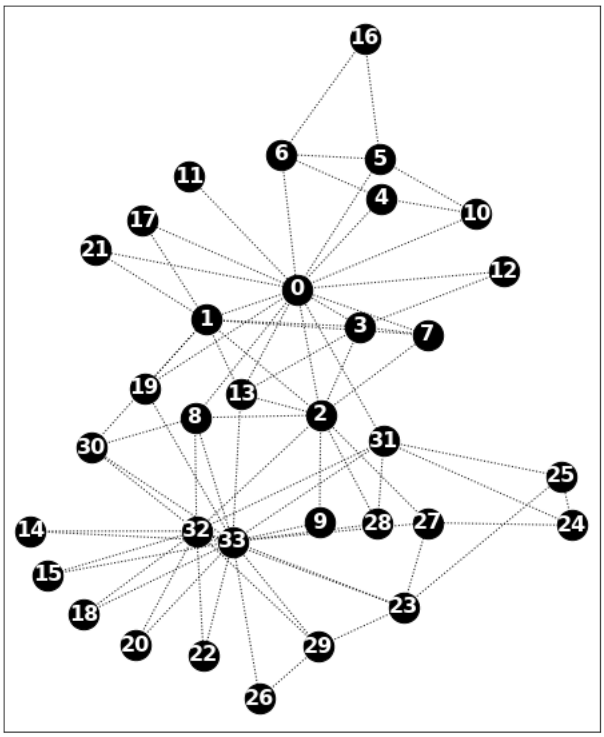
To work similarly as (GOLDSHTEIN et al., 2004) we define it similar to the global efficiency as seen in (LATORA; MARCHIORI, 2001) in Equation 3.2

$$Com(\mathbf{G}) = \frac{1}{N(N-1)} \sum_{\mathbf{G}} Com_{ij}. \tag{3.7}$$

Where $Com(G)$ is the global communicability of the network G, Com_{ij} is the communicability between the nodes i and j, and N is the number of nodes.

This algorithm will be tested in the Zachary's Karate Club network.

Figure 3.2 - Zachary's Karate Club graph, used to test our results.



Source: Zachary (1977).

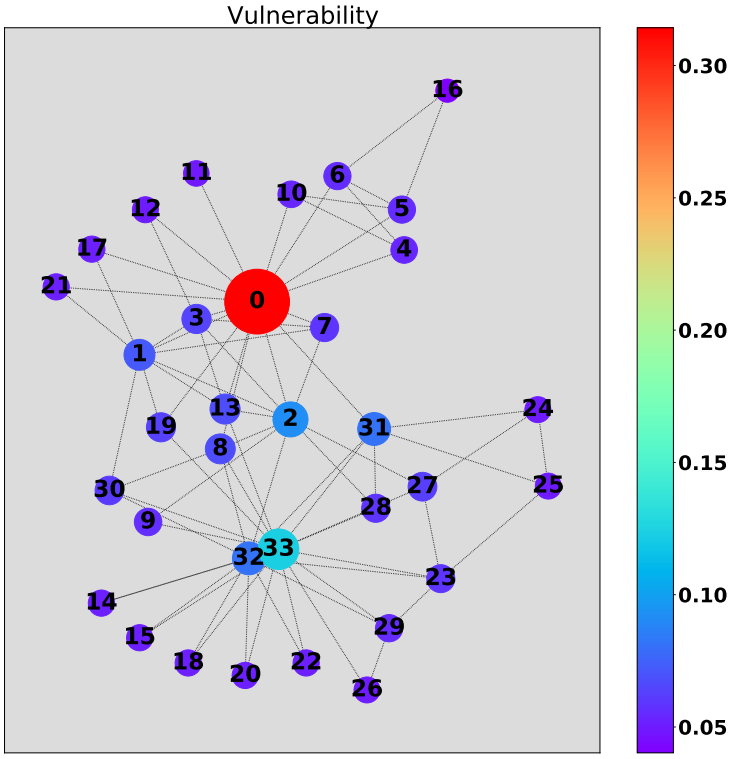
The Zachary's Karate Club is a well-known social network used to study fission or faction formation. It is a university-based karate club, in which factional fission led to a formal separation of the club into two organizations. The karate club was observed for three years, from 1970 to 1972, and the study was published in 1977. At the beginning of the study, there was an incipient conflict between the club

president, John A., and Mr. Hi over the price of karate Lessons. Mr. Hi, who wished to raise prices, claimed the authority to set his lesson fees, but John A., who wished to stabilize prices, claimed the authority since he was the club's chief administrator (ZACHARY, 1977). As time went by, everyone got involved, and the fission of the club happened; Mr. Hi formed a new organization taking some students with him. In the graph, Mr. Hi is represented by node 0 and John A. by node 33.

3.4 Results

In this Section, it becomes clear the differences between both vulnerabilities. First, we show the results to the one using efficiency as performance.

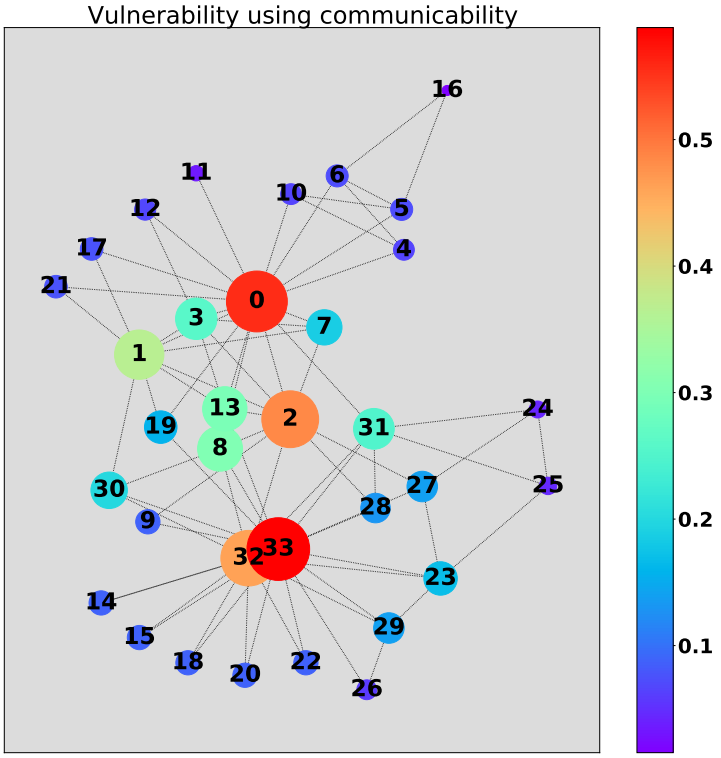
Figure 3.3 - Visualizing the point-wise vulnerability of the Zachary Karate Club. This vulnerability is calculated using efficiency as performance.



As we can see in Figure 3.3, there is only one highly vulnerable node. The node 0 carries the most significant value of 0.314, followed by the node number 33 with a value of 0.123, proving that those two nodes are of utmost importance to the efficiency of the system, and removing them will cause the efficiency to drop drastically in the graph. The drop in efficiency happens because the graph's shortest paths mostly go through those two nodes.

Now, when we use communicability as performance, we are analyzing something different than shortest paths.

Figure 3.4 - A visualization of the point-wise vulnerability of the Zachary Karate Club. This vulnerability is calculated using communicability as performance.



Which becomes clear in Figure 3.4. Not only the nodes 0 and 33 are present as points of vulnerability, but some others begin to take priority. The node 0 loses its position as the most important node to node 33, meaning that many other walks go through it when we compare to the shortest paths going through node 0.

Table 3.1 - Table relating the values of vulnerability with two different performances, being them efficiency Vul(Efi), and communicability Vul(Com). In the last row we can see the standard deviation of each vulnerabilities.

Node	Vul(Efi)	Vul(Com)	Node	Vul(Efi)	Vul(Com)
0	0.314	0.555	17	0.051	0.077
1	0.072	0.365	18	0.051	0.088
2	0.092	0.484	19	0.063	0.158
3	0.064	0.257	20	0.051	0.088
4	0.053	0.065	21	0.051	0.077
5	0.056	0.073	22	0.051	0.088
6	0.056	0.073	23	0.059	0.168
7	0.059	0.187	24	0.050	0.043
8	0.067	0.303	25	0.050	0.046
9	0.056	0.089	26	0.051	0.059
10	0.053	0.065	27	0.062	0.139
11	0.049	0.035	28	0.059	0.131
12	0.051	0.067	29	0.056	0.140
13	0.068	0.295	30	0.061	0.199
14	0.051	0.088	31	0.081	0.248
15	0.051	0.088	32	0.081	0.463
16	0.040	0.015	33	0.123	0.589
Range(Vul(Efi)):	0.04	0.314	Range(Vul(Com)):	0.015	0.589
std(Vul(Efi)):	0.045		std(Vul(Com)):	0.153	

In the table, we highlight the most significant differences in both metrics. As we can see, both of them highlight different types of vulnerabilities in the graph and can be used in different cases and scenarios.

Another notable difference, in this specific case, is the range of both vulnerabilities. The standard deviation for each vulnerability is 0.045, using Efficiency as performance, and 0.152 using Communicability as performance. A broader range indicates that a node can have a heavier influence than others.

It is curious to note that both nodes 0 and 33, representing Mr. Hi and John A., have the most significant value of vulnerability in both cases. However, vulnerability using Communicability highlights better the importance of those nodes.

3.5 Conclusion

In this work, we reach the definition of a vulnerability variant using Communicability as performance and successfully compare it to its counterpart of efficiency as

performance.

Here we show the broad difference between them in highlighting different nodes as vulnerable given efficiency uses only shortest paths and Communicability goes beyond and uses all paths that go through a node. The difference of the vulnerabilities standard deviations is 0.108, telling us that the Vulnerability with Communicability has a broader range than the Vulnerability with Efficiency. The Vulnerability with Communicability also shows more highly vulnerable nodes, with a total of 4 nodes with its vulnerability more greater than 0.4, while the Vulnerability with Efficiency only highlights 1 node as highly vulnerable.

It is important to note how the Vulnerability with Communicability successfully identified the core members from the Karate Club, the master, and the administrator. When the fission happened, the club was divided within them.

In conclusion, both metrics present a distinct point of view inside the same network and are meant to be used in different approaches, the context in which those metrics are applied matters when we have to choose between them. The communicability computational cost cannot be disregarded when applied to more extensive networks - like a road map or an electric grid - the matrix multiplication used in the process is hard to deal with without the help of high-performance computing. However, we are already working with this and achieving good results.

4 VULNERABILITIES AND HOW THEY RELATE WITH OTHER METRICS

The previous Chapter left us with a task to validate the proposed metric of Vulnerability with Communicability. This Chapter is an effort to investigate this by exploring the relations of Vulnerabilities with other metrics and between themselves. Those differences depend on the network's topology, so we select the two models, Erdős–Rényi, and Barabási-Albert, to study the previously cited relations.

4.1 Introduction

There are plenty of metrics when talking about Complex Networks (COSTA et al., 2007), and each of them is applied to measure a specific quality of the network (LATORA; MARCHIORI, 2001; LATORA; MARCHIORI, 2004; GOLDSHTEIN et al., 2004; ESTRADA; HATANO, 2008). However, those metrics may have a relation between them, which may vary depending on the model.

In this work, we already defined a new type of metric - in Chapter 3. This metric is a twist in an already published metric of Vulnerability. Instead of using efficiency as a performance, we use communicability to understand how the diffusion inside the network can differ in vulnerability points, considering not only the shortest path between each pair of nodes.

This Chapter relates different metrics already established in the literature with two distinct vulnerability indexes. We explain our methodology to relate the metrics shown in Chapter 2 and show their relations in Section 4.3. Finally, a conclusion is presented at Section 4.4.

4.2 Methodology

Our methodology begins by generating similar graphs and keeping an approximate ratio of nodes and edges. To generate, we use two different models, Erdős–Rényi (Random Network) and Barabasi-Albert (Scale-free Network).

To both models, we generate four graphs while changing parameters to understand how the relation between metrics can change. The four cases are $N = 100$ and $L \approx 1400$, $N = 100$ and $L \approx 600$, $N = 256$ and $L \approx 1400$, $N = 2000$ and $L \approx 100000$.

For each graph, we calculate the following metrics:

- Degree
- Betweenness
- Mean Shortest Path
- Vulnerability with Efficiency
- Vulnerability with Communicability

Then, we make a scatter plot comparing each metric. We also fit a curve with the parameters and display the R^2 associated. The coefficient of determination (or R^2) is defined as

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}. \tag{4.1}$$

Where SS_{res} is the residual sum of squares, and SS_{tot} is the total sum of squares. They are both defined using the value of the observations against the data resulting in the fit. Mathematically speaking, they are

$$SS_{res} = \sum_i (y_i - f_i)^2, \tag{4.2}$$

and

$$SS_{tot} = \sum_i (y_i - \bar{y})^2. \tag{4.3}$$

Where y_i is our observation, in our case, it is the metrics of Degree, Betweenness, and Mean Shortest Path. $\bar{y}$ is the mean of y_i, f_i is the calculated y_i obtained in the fit.

This work give us an insight into how the density of the graph can change the relations between metrics, so we explore this idea by making other plots increasing the density of the graph by generating a graph with more edges and the same number of nodes. Then, we calculate the metrics and make the relation only to calculate the R^2, which is kept and put in a plot versus the graph density.

We executed the program in UFOP's cluster, an AMD Ryzen Threadripper 3960X

with 24 cores (48 threads) and 3.70GHz, 128GB RAM DDR4. GPU RTX 3090, with 24GB RAM GDDR6X and more than 10 thousand Cuda cores.

4.3 Results

We show the results in a series of scatter plots, relating the vulnerabilities with the other three metrics and with each other. We repeat the comparison to each different graph model of Erdős–Rényi, and Barabasi-Albert. This way, we can observe how different kinds of networks change the relations between the metrics.

The following Figures show the results arranged in a table-like form. The first line compares Vulnerability with Efficiency as performance with the degree, betweenness, and mean shortest path in a scatter plot. The second line does the same thing to Vulnerability with Communicability as performance.

After that, we bring up distributions for each graph made in this Chapter.

Lastly, we display the graph density versus the correlation of fits, divided by the types of vulnerabilities and the graph's number of nodes.

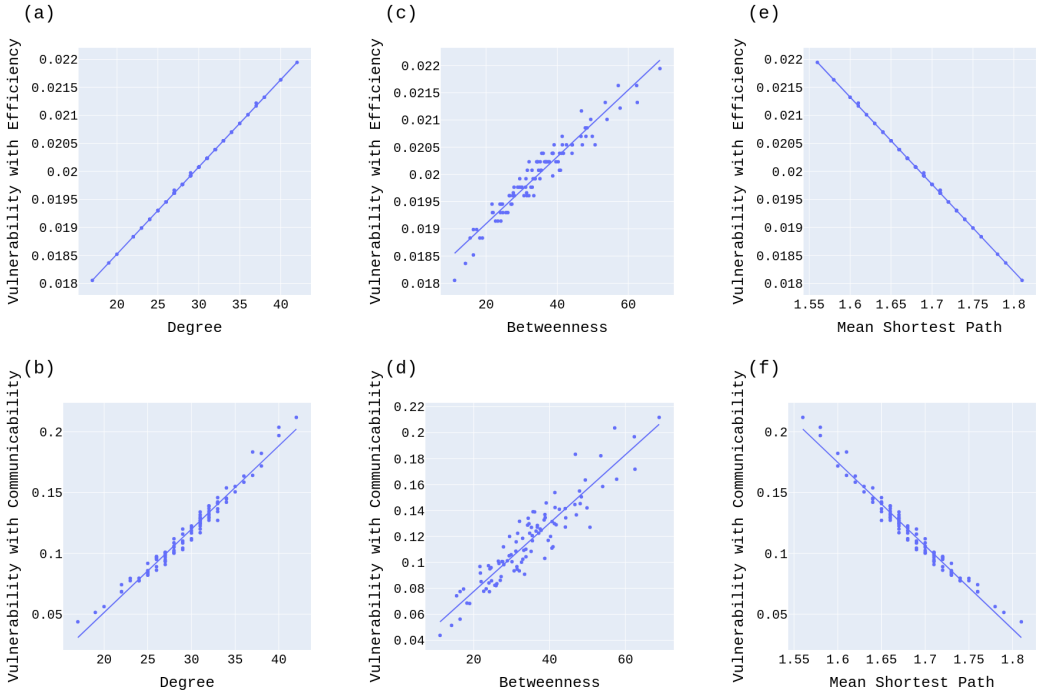
4.3.1 Erdős–Rényi model

We start our result section by showcasing the Erdős–Rényi model, with a set of plots and comparisons between metrics.

4.3.1.1 Scatter plots

Firstly, we introduce the scatter plots to each different graph. We start with our reference case of $N = 100$ and $L = 1475$. Forward, we change the number of nodes and edges to explore the differences that can arise.

Figure 4.1 - Scatter plots to graph with following characteristics: N = 100, L = 1475 d = 0.30, $< c > = 0.30$, $< k > = 29.50$, D = 2, $< l > = 1.68$.



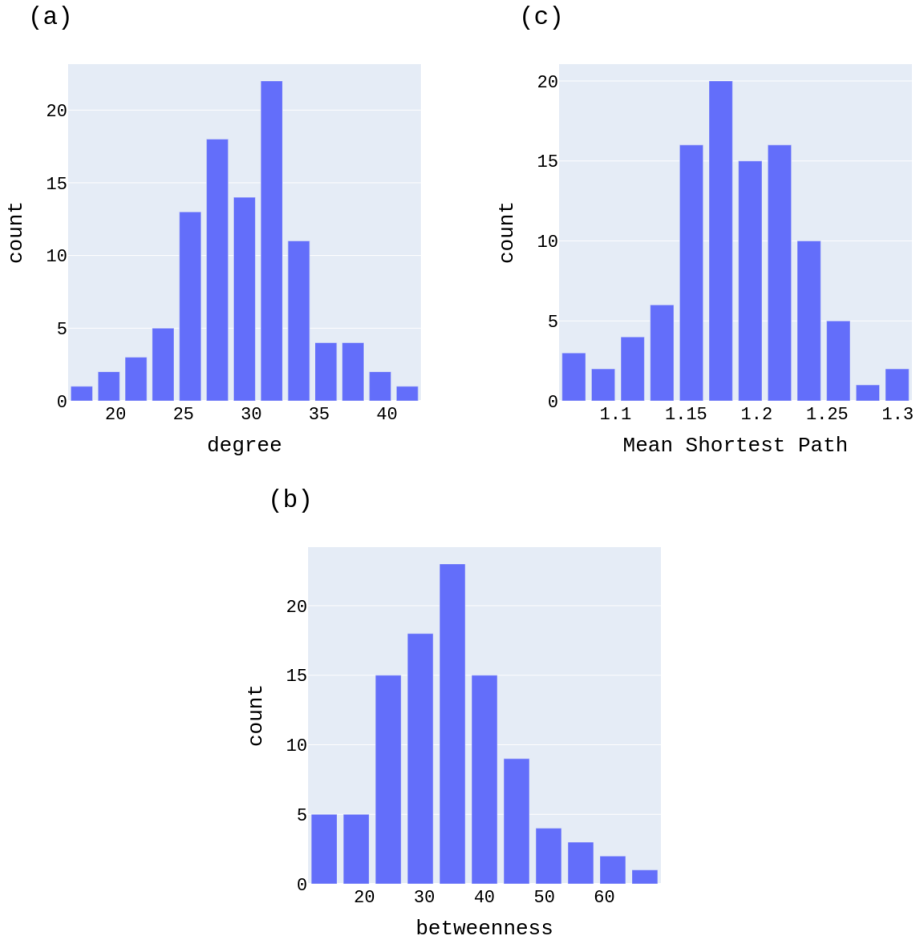
In our first case in Figure 4.1, we can see a strong linear relation between all metrics. Each vulnerability increase with the degree and betweenness and decreases with the length of the mean shortest path associated with the node.

Table 4.1 - Table associated with Figure 4.1.

Image	(a)	(b)	(c)	(d)	(e)	(f)
R^2	1	0.971	0.930	0.858	1	0.971

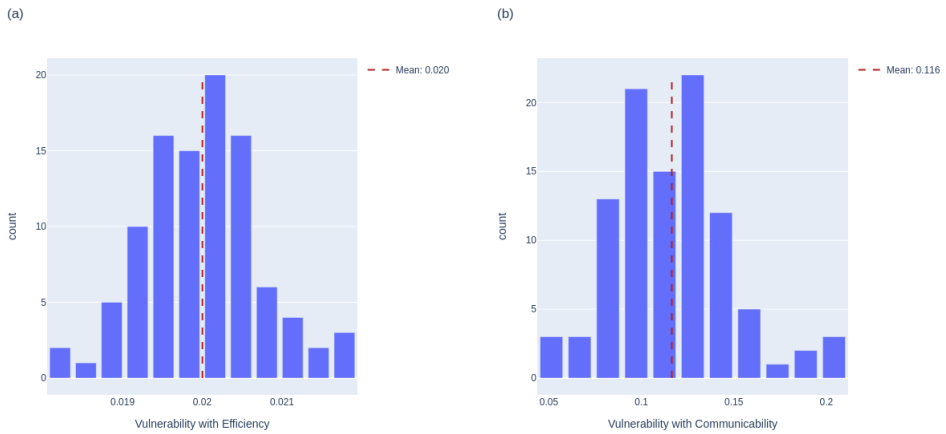
In Table 4.1, we see the coefficient of determination reaching the maximum value of 1 and the lowest of 0.858.

Figure 4.2 - Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Figures 2.4, 2.5.



In Figure 4.2 we see all distributions in a bell-like shape, which is expected to the degree in the Erdős–Rényi model.

Figure 4.3 - Erdős–Rényi distribution of each vulnerability shows how many times a value inside the bin's interval appears. The line represents the sample's mean value.

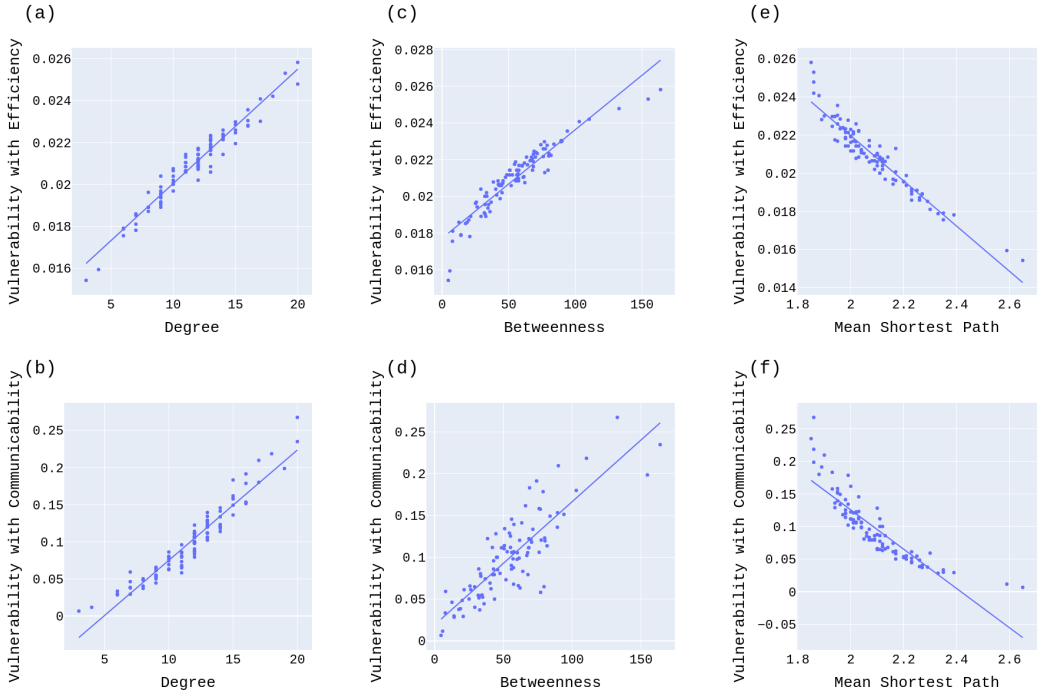


The vulnerabilities in Figure 4.3 present the same bell-like shape. This behavior of Vulnerability is often associated with being more vulnerable to random failures on the network.

Those distributions are similar to all cases, so we show them only for our reference case, and all others can be found in Appendix A.

Back to the scatter plots, we decrease the number of edges to $L = 584$, maintaining the same number of nodes $N = 100$; this way, we achieve a less-connected graph.

Figure 4.4 - Scatter plots to graph with following characteristics: N = 100, L = 584, d = 0.11, $< c > = 0.11$, $< k > = 11.68$, D = 4, $< l > = 2.08$.



In Figure 4.4 we notice how the drop in density affects the scatter plots appearance. They are still similar, with strong linearity; however, a sensible change can be noticed in the distribution of points.

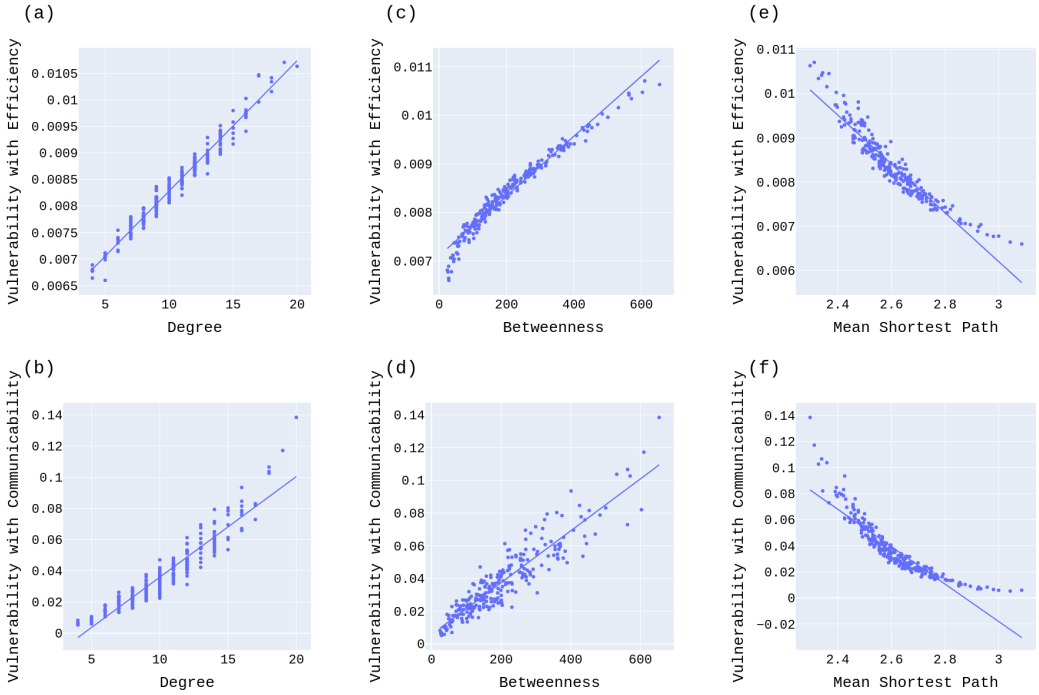
Table 4.2 - Table associated with Figure 4.4.

Image	(a)	(b)	(c)	(d)	(e)	(f)
R^2	0.945	0.915	0.883	0.715	0.903	0.760

Table 4.2 represents well that change. While in Table 4.1 the maximum value reached 1, Table 4.2 maximum value is 0.945 and minimum of 0.715. This indicates that graph density is linked to the linearity of those relations.

Now, we increase the number of nodes and keep a similar edges count, changing to $N = 256$ and $L = 1329$: resulting in a drop of density, going from density $d = 0.3$ (in our reference case) to $d = 0.04$ (in this one).

Figure 4.5 - Scatter plots to graph with following characteristics: N = 256, L = 1329, d = 0.04, $< c > = 0.04$, $< k > = 10.38$, D = 4, $< l > = 2.61$.



The drop in density changes the relation between the metrics, as we show in Figure 4.5. They behave the same way; a higher degree and betweenness increase each Vulnerability. At the same time, a more significant Mean Shortest Path decreases both the vulnerabilities.

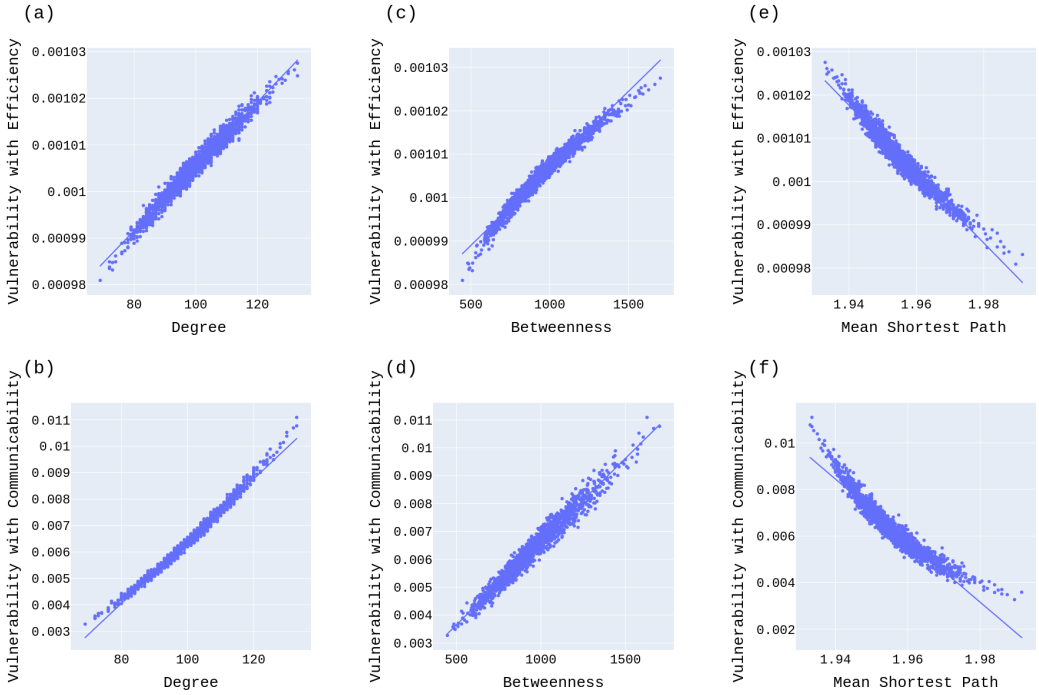
Table 4.3 - Table associated with Figure 4.5.

Image	(a)	(b)	(c)	(d)	(e)	(f)
R^2	0.968	0.893	0.957	0.847	0.899	0.810

The associated R^2 as shown in Table 4.3 indicates to us that even with the drop in density, the linearity of some relations can still increase with the increase of total nodes.

So, to investigate, we generated a graph with $N = 2000$ and $L = 99685$ and a similar density as the previous case of $d = 0.05$.

Figure 4.6 - Scatter plots to graph with following characteristics: N = 2000, L = 99685, d = 0.05, $< c > = 0.05$, $< k > = 99.68$, D = 3, $< l > = 1.95$.



Then, in Figure 4.6 we observe how this linearity is better represented in a graph with more nodes. They behave the same way as all previous images.

Table 4.4 - Table associated with Figure 4.6.

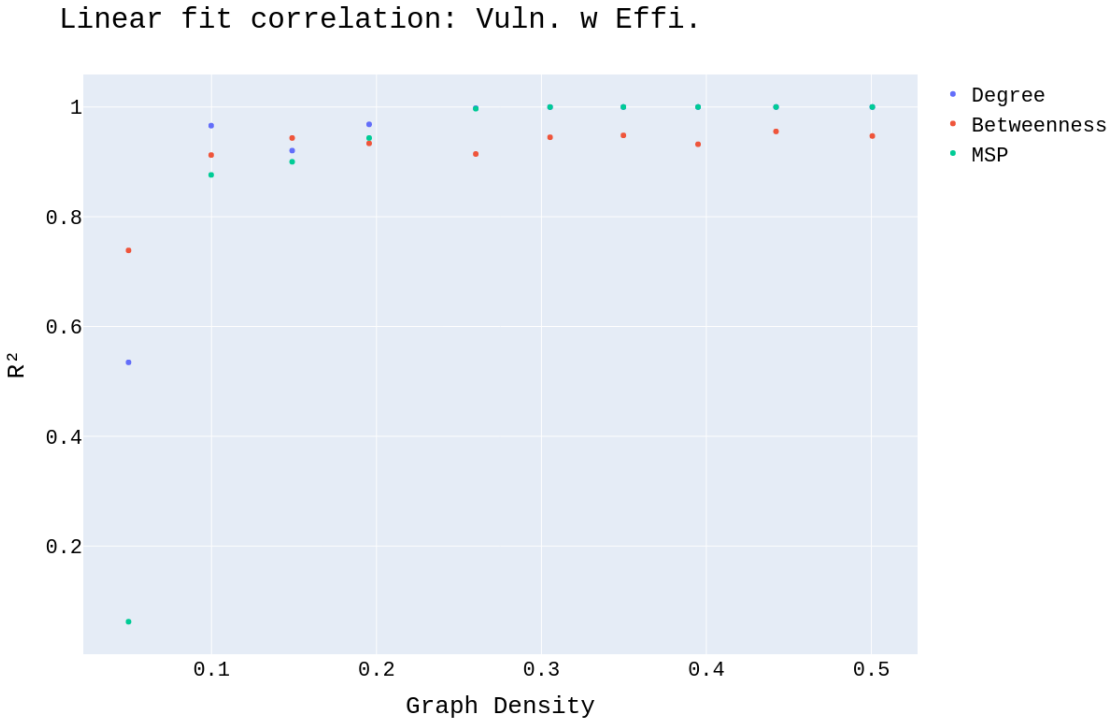
Image	(a)	(b)	(c)	(d)	(e)	(f)
R^2	0.967	0.988	0.968	0.963	0.951	0.919

It is interesting to note how the R^2 value reflects the increase of linearity as the number of nodes and edges increases, maintaining a low density.

4.3.1.2 Density variation

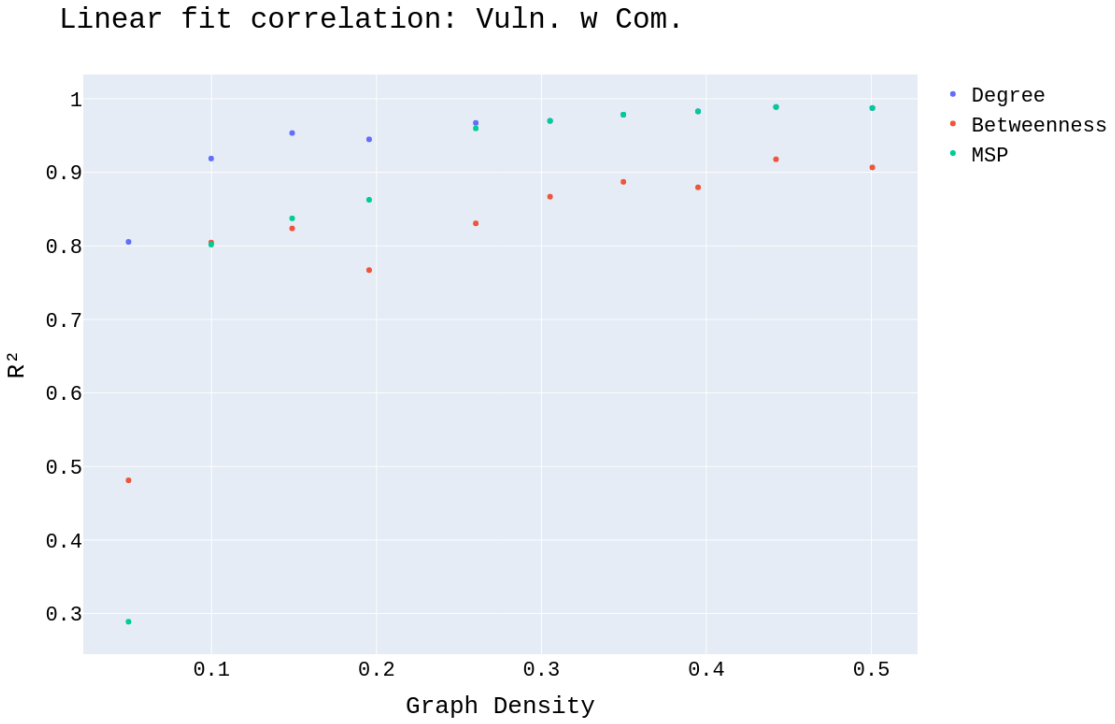
This type of analysis shows how the density is related to the R^2 value. So, we explore two types of graphs, with $N = 100$ and $N = 256$, using both models presented in this work. The following Figures relate the value of R^2 in a linear fit and graph density when comparing the metrics of Degree, Betweenness, and Mean Shortest Path with Vulnerability with Efficiency and Vulnerability with Communicability.

Figure 4.7 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$.



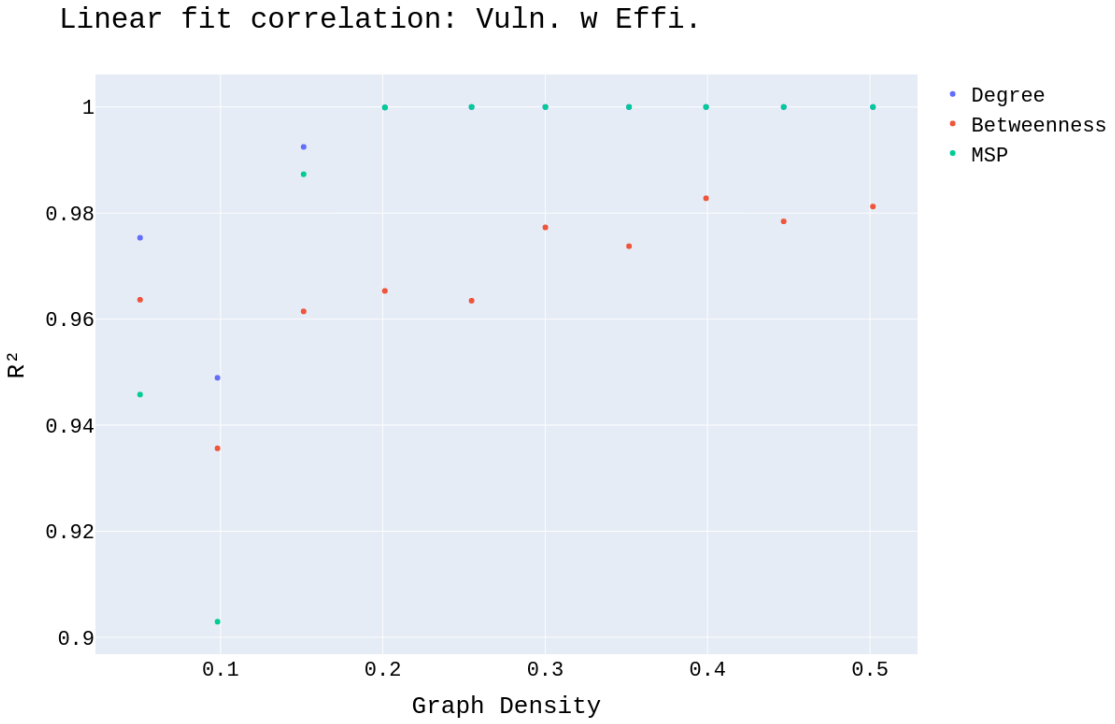
We notice in Figure 4.7 how the scatter between Vulnerability with Efficiency and the Mean Shortest Path has a low linear fit R^2 to low-density graphs, while Betweenness and Degree has a more significant value. Nonetheless, the coefficient of determination value rapidly increases with the graph density, reaching the limit of 1.

Figure 4.8 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$.



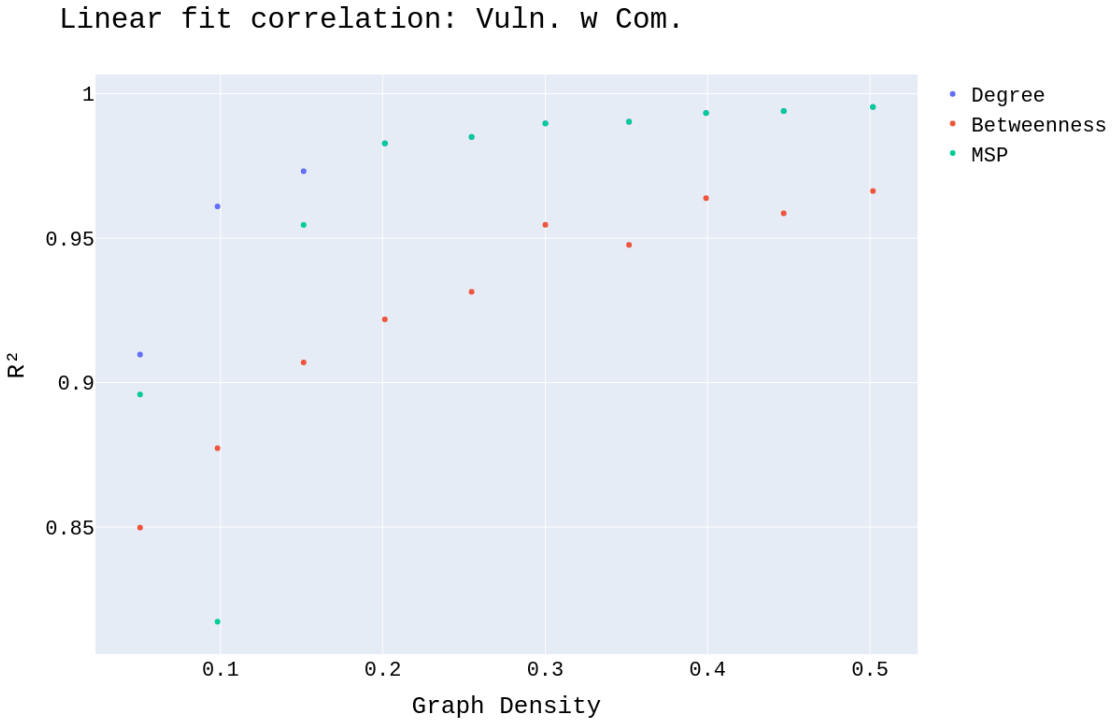
In Figure 4.8 we notice a similar pattern, with the Mean Shortest Path quickly taking the lead and approaching the value of 1 when testing linearity via the R^2 score.

Figure 4.9 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$.



With Figure 4.9 we notice a discrepancy in the relations with the Vulnerability with Efficiency, where the R^2 does not steadily increase but instead has a drop in graphs with $d \approx 0.1$ and then increases. However, it is worth to mention that the range in y-axis goes from 0.9 to 1, meaning a small variation in the result.

Figure 4.10 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$.



The previous behavior of Figure 4.9 is not identified in Figure 4.10 to all metrics. Instead, the R^2 only decreases at $d = 0.1$ in the Mean Shortest Path relation with the Vulnerability with Communicability.

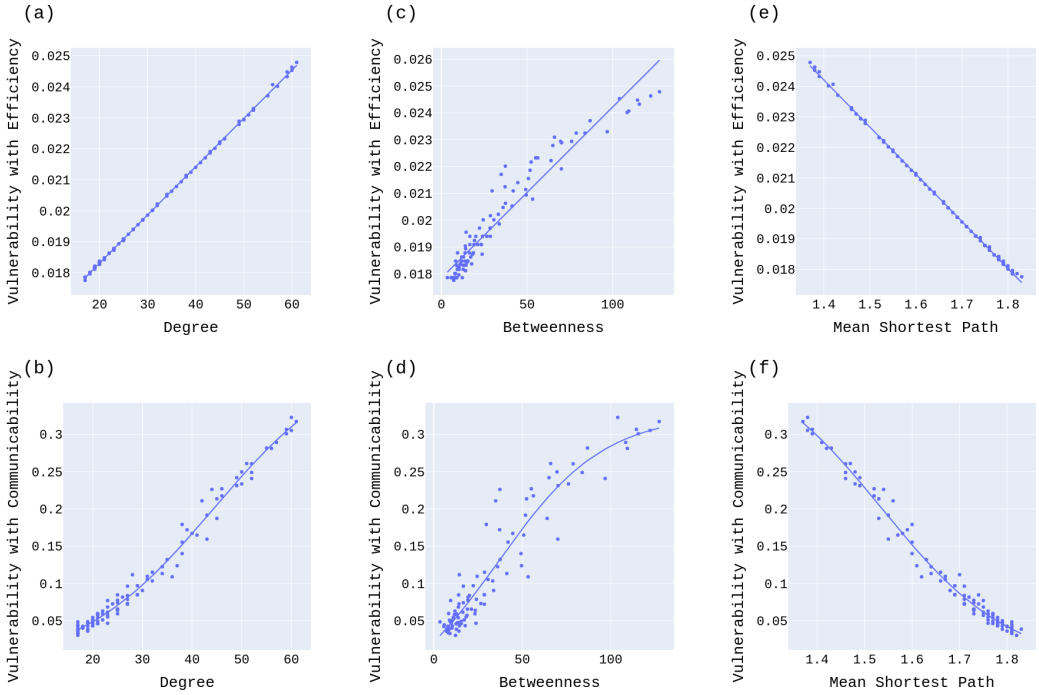
4.3.2 Barabási-Albert model

Now, we go through the Barabási-Albert model; these results are slightly different from the previous.

4.3.2.1 Scatter plots

In the same way as before, we start with the scatter plots of our reference case, with $N = 100$ and $L = 1547$.

Figure 4.11 - Scatter plots to graph with following characteristics: N = 100, L = 1547, d = 0.3, $< c > = 0.42$, $< k > = 30.94$, D = 3, $< l > = 1.67$.

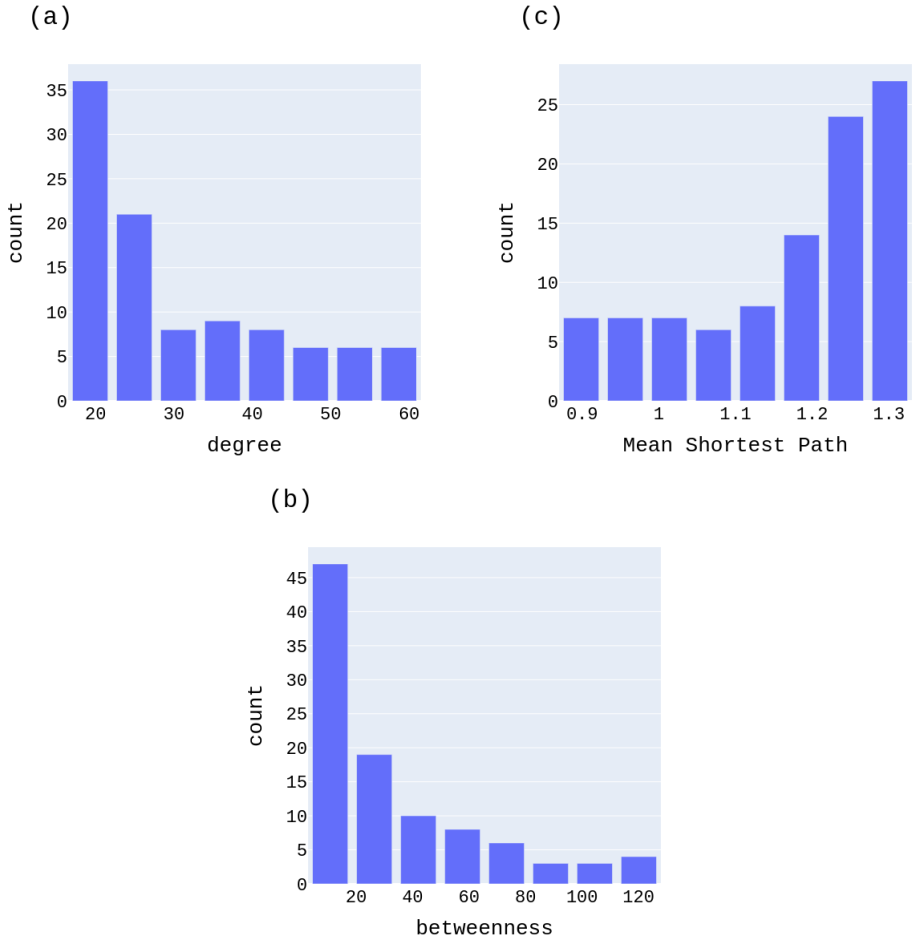


Unlike the Erdős–Rényi model, the Barabási-Albert shows not only linear relations, but sigmoid as well. In Figure 4.11 we can see in some places how the sigmoid function fits better than linear. Besides the change in relation, there is always an increase in Vulnerability when the degree and betweenness increase, while the mean shortest path length increasing leads to a decrease in Vulnerability.

Image	(a)	(b)	(c)	(d)	(e)	(f)
R^2	0.977	0.986	0.926	0.906	0.999	0.877
Type	Linear	Sigmoid	Linear	Sigmoid	Linear	Sigmoid

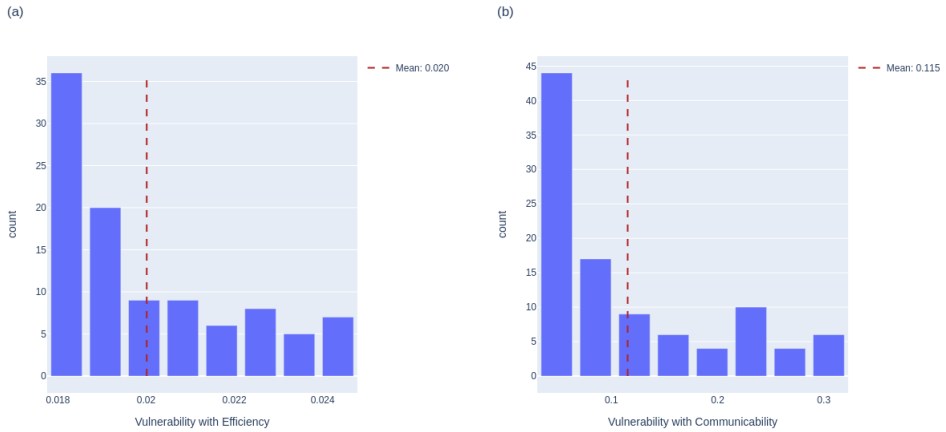
Table 4.3.2.1 shows the quality of the relation between them. The R^2 minimum is 0.877 to the sigmoid function, showing the quality of our fits. The maximum value reached is 0.999, which almost represents a perfect fit.

Figure 4.12 - Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Figures 2.5, 2.5.



In Figure 4.12 we can see the difference in distributions from Figure 4.2. Here, all cases follow a power-law-like distribution, in a decreasing way in (a) and (b) and increasing in (c). This is a Scale-free model characteristic to the degree, leading to a graph with more nodes with a lower degree.

Figure 4.13 - Barabási-Albert distribution of each vulnerability shows how many times a value inside the bin's interval appears. The line represents the sample mean value.

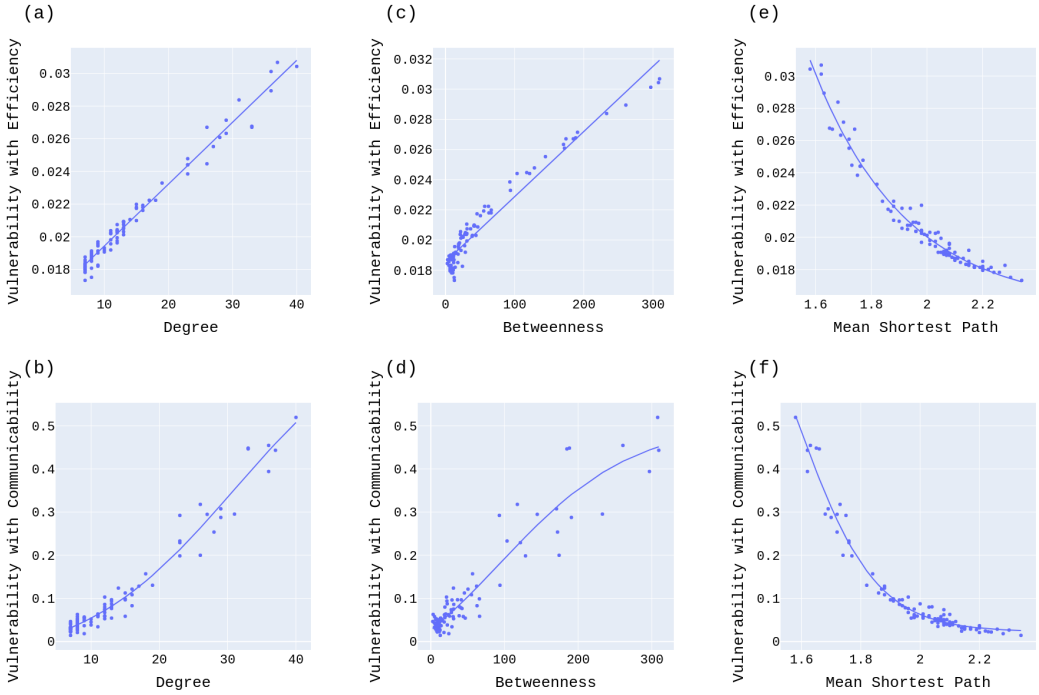


The Vulnerability reflects this difference in distributions, showing a power-law-like distribution in Figure 4.13. Those networks are more susceptible to target attacks - when reaching the hubs.

It is interesting to note how each vulnerability has a different maximum value between efficiency and communicability, and also when comparing the models of Erdős–Rényi and Barabási-Albert.

We explore other configurations, generating graphs with more nodes and fewer edges, as we did with the Erdős–Rényi model.

Figure 4.14 - Scatter plots to graph with following characteristics: N = 100, L = 672, d = 0.13, $< c > = 0.23$, $< k > = 13.44$, D = 3, $< l > = 1.99$.



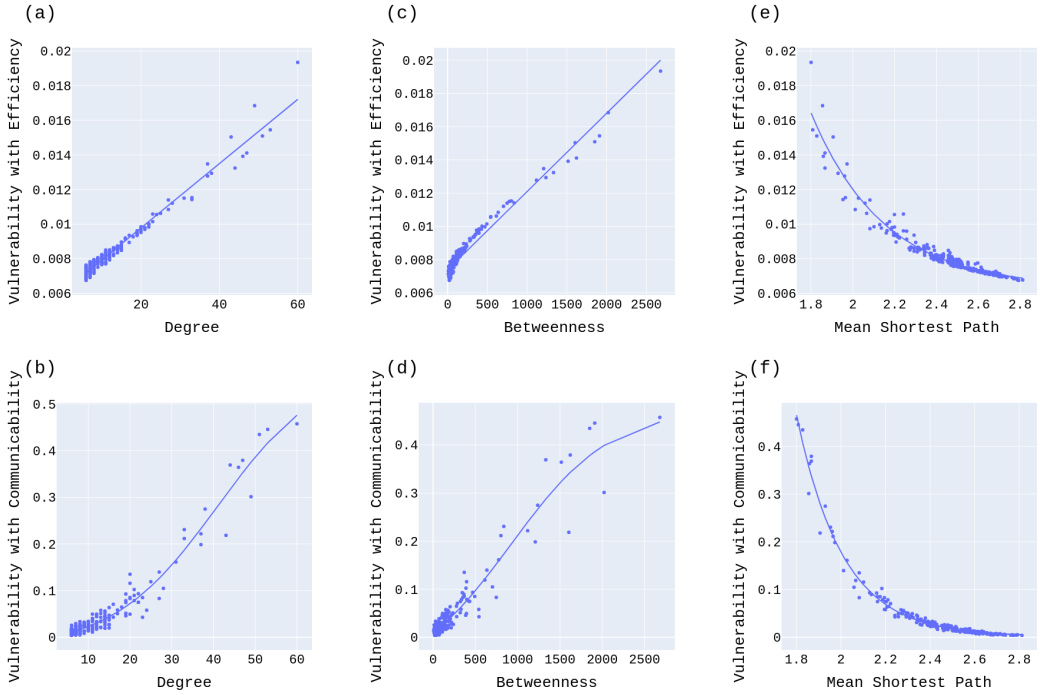
With the drop in the number of edges, we can see the difference in each image in Figure 4.14. We change the image (e) to a sigmoid fit to better reflect the disposition of points. There is still an increase in Vulnerability with the degree and betweenness, while it decreases with the size of the mean shortest path.

Image	(a)	(b)	(c)	(d)	(e)	(f)
R^2	0.901	0.965	0.950	0.901	0.773	0.891
Type	Linear	Sigmoid	Linear	Sigmoid	Sigmoid	Sigmoid

The coefficient of determination persists with a high value, as we can see in 4.3.2.1, close to one. Even when changing the fit to sigmoid in the image (e), the R^2 value is smaller, showing the difficulty of the fitting.

We increase the number of nodes in the generated graph to $N = 256$ and edges to $L = 1515$.

Figure 4.15 - Scatter plots to graph with following characteristics: N = 256, L = 1515, d = 0.05, $< c > = 0.10$, $< k > = 11.83$, D = 4, $< l > = 2.45$.



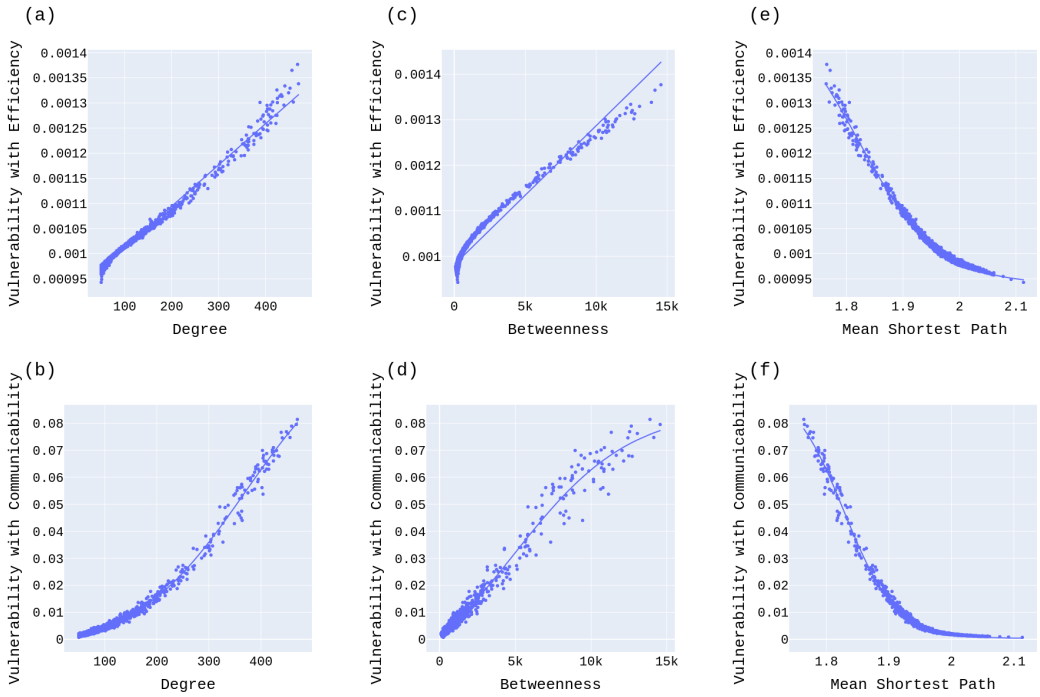
In Figure 4.15 we see the same behavior as before; the vulnerabilities increase with the degree and betweenness while decreasing with the mean shortest path length.

Image	(a)	(b)	(c)	(d)	(e)	(f)
R^2	0.879	0.957	0.958	0.931	0.793	0.893
Type	Linear	Sigmoid	Linear	Sigmoid	Sigmoid	Sigmoid

Looking at the coefficient of determination to the Barabási-Albert model in Table 4.3.2.1, we notice how this model behaves differently to the drop of density. Here, the R^2 most significant change is for the relation between the degree with Vulnerability with Efficiency, while the Erdős–Rényi's most significant drop is for the relation between the mean shortest path with Vulnerability with communicability.

Now, we increase the number of nodes and edges and keep the density low, changing to $N = 2000$ and $L = 98725$, with $d = 0.05$.

Figure 4.16 - Scatter plots to graph with following characteristics: N = 2000, L = 98725, d = 0.05, $< c > = 0.12$, $< k > = 98.72$, D = 3, $< l > = 1.97$.



In Figure 4.16, we notice how each image takes a sharper shape, with points being closer together, even with a low density.

Image	(a)	(b)	(c)	(d)	(e)	(f)
R^2	0.990	0.993	0.986	0.981	0.948	0.966
Type	Linear	Sigmoid	Linear	Sigmoid	Sigmoid	Sigmoid

This behavior is reflected at Table 4.3.2.1, with R^2 values all closer to 1.

4.3.2.2 Density variation

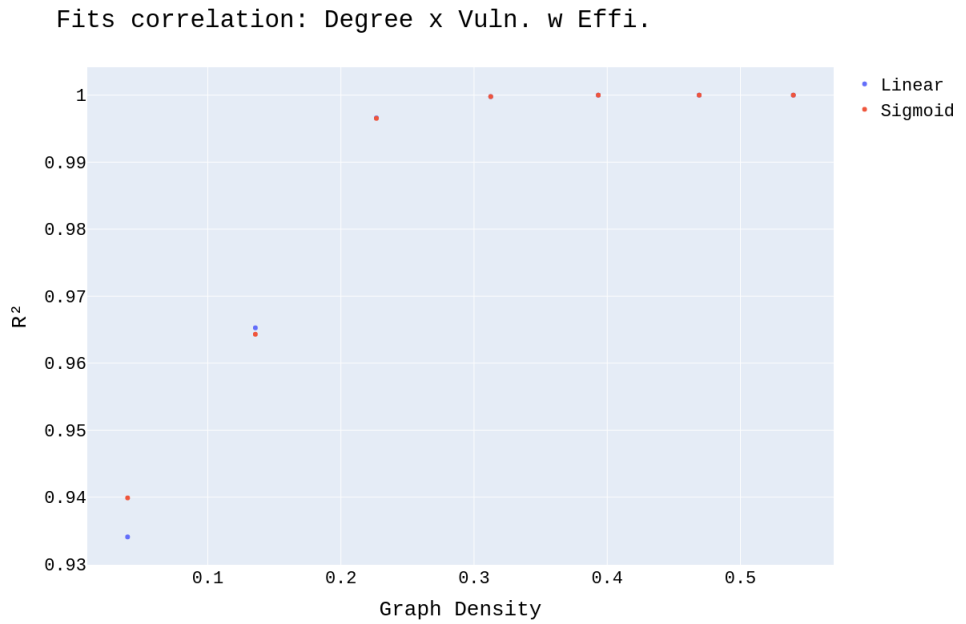
In the same way as before, we notice the changes in R^2 with density. We also explore this concept in the Barabási-Albert model. Since we do two types of fits in this part, we explore them in separate Figures for each relation.

We calculate the R^2 to each relation between metrics, with densities going from 0.1 to 0.5. This proceeding is repeated to the linear and sigmoid functions, helping us

decide which one to use.

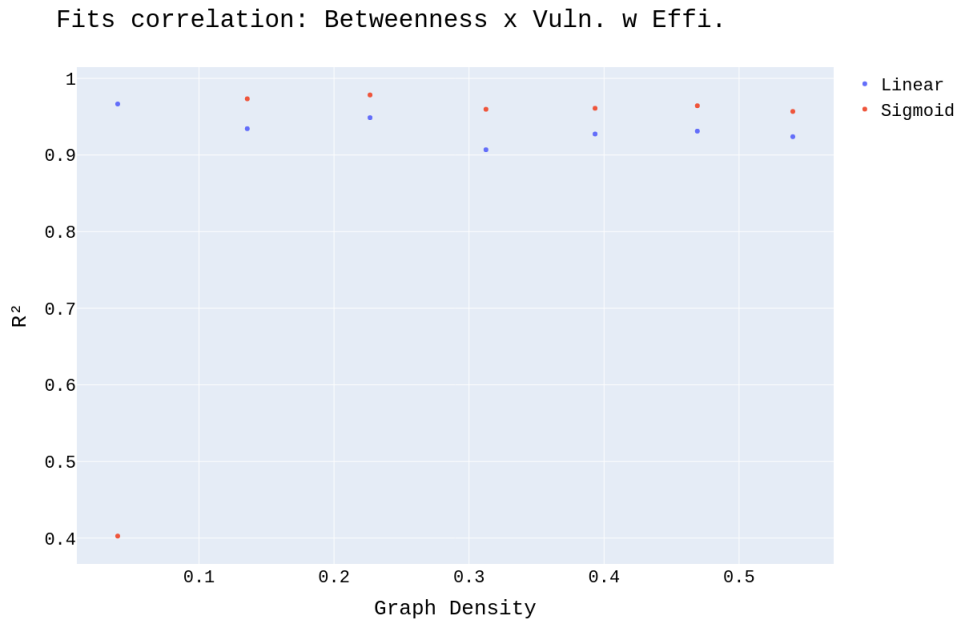
The entire process is repeated to graphs with 100 and 256 nodes.

Figure 4.17 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$.



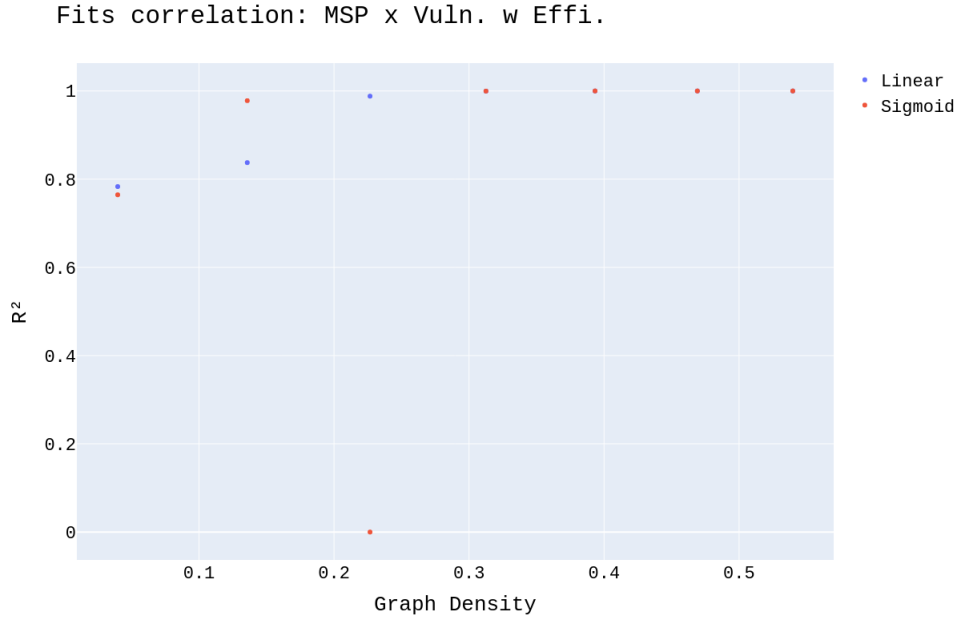
First, we see the relation between degree and Vulnerability with efficiency in Figure 4.17. Both R^2 increase with graph density, and since the linear function is more straightforward than sigmoid, we tend to choose it.

Figure 4.18 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$.



In Figure 4.18 we see a lower variation of R^2 when comparing the betweenness with the Vulnerability with efficiency. With a minimum value of 0.91 and a maximum of 0.98, the sigmoid function presents better results, but it is such a low difference that we tend to choose the linear function since it is straightforward.

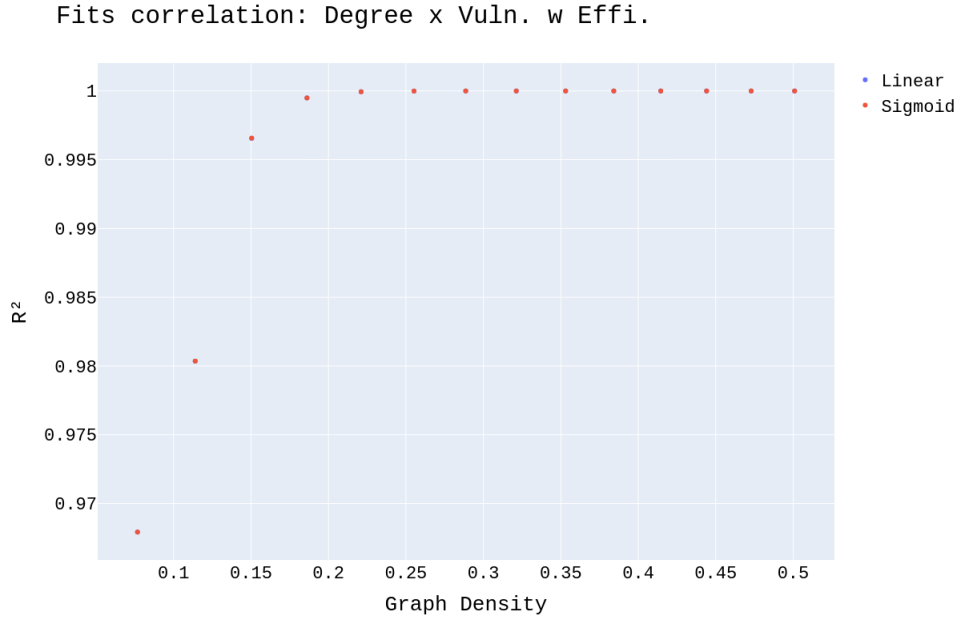
Figure 4.19 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$.



Meanwhile, in comparison with the mean shortest path length in Figure 4.19, we see the sigmoid function working better in all cases over 0.3 density. Nonetheless, the sigmoid fit over the points is visually better than the linear fit, so we choose it.

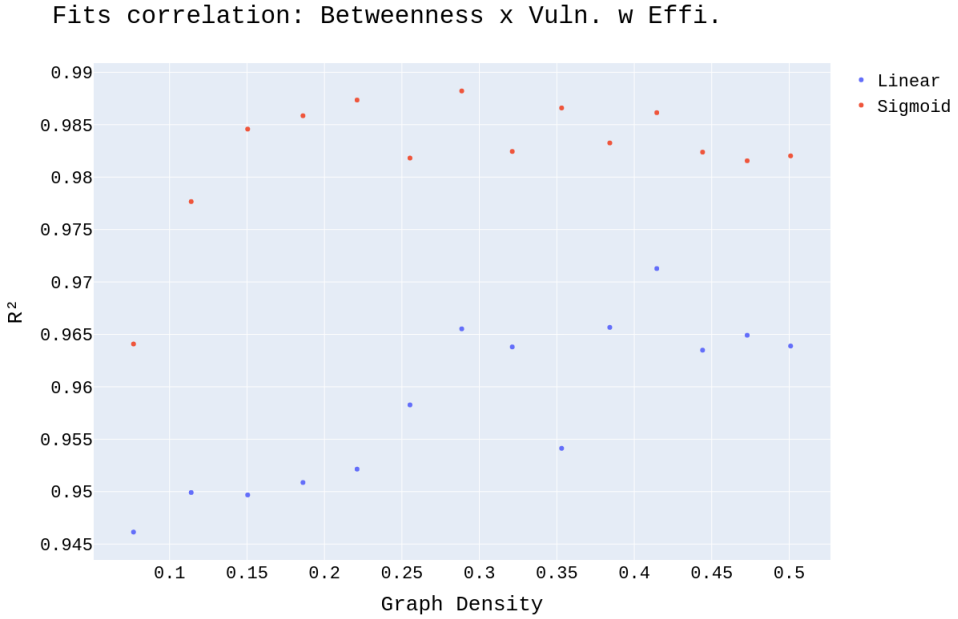
We increase the generated graph's size to $N = 256$ and repeat the same graphs.

Figure 4.20 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$.



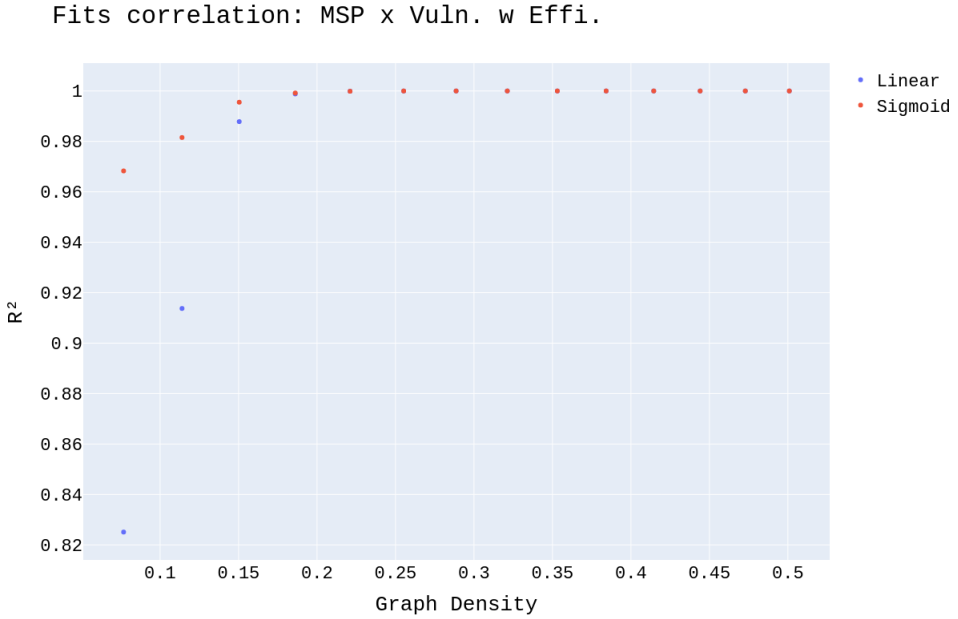
In Figure 4.20 we can see that the relation of the degree with the vulnerability with efficiency maintains the same increase in R^2 as Figure 4.17. Both Figures show a close result between both fits and converging at 1.

Figure 4.21 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$.



Now looking at the betweenness in Figure 4.21, we can see how close the points are, with the minimum value around 0.94 and a maximum value of 0.98. With the disposition of points, this result leads us to choose a linear fit to the relation since it is easier.

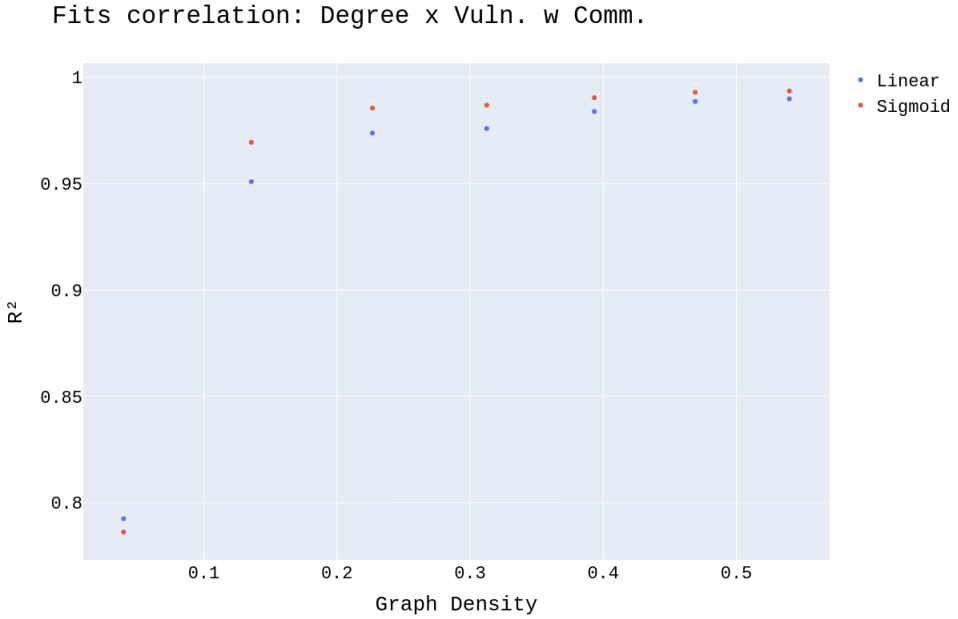
Figure 4.22 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$.



To the mean shortest path in Figure 4.22, the analysis persists the same as Figure 4.19, where the sigmoid is usually better than the linear fit by little, converging at 1 as well.

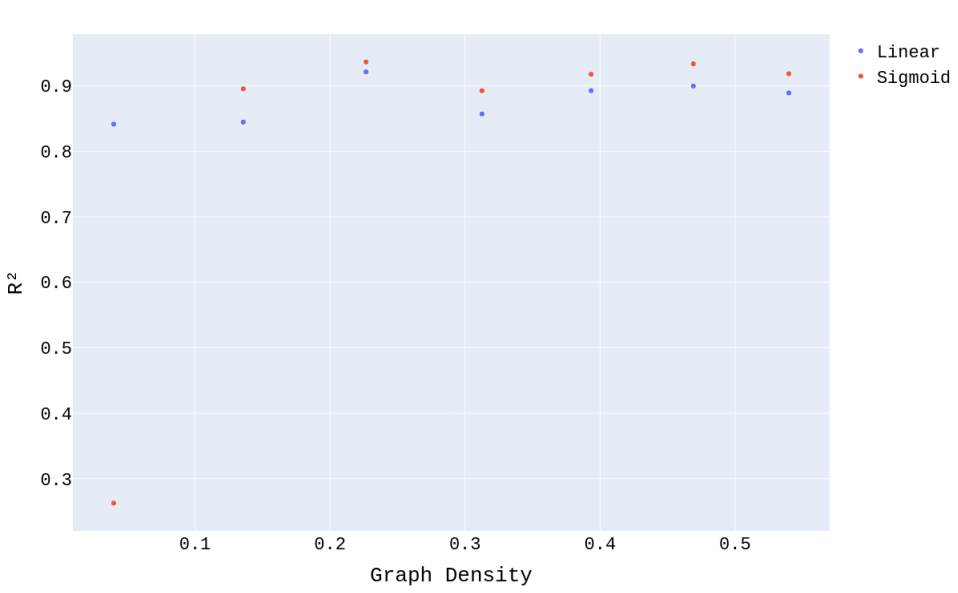
Now, we analyze the relations between Vulnerability with Communicability, starting with the degree.

Figure 4.23 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$.



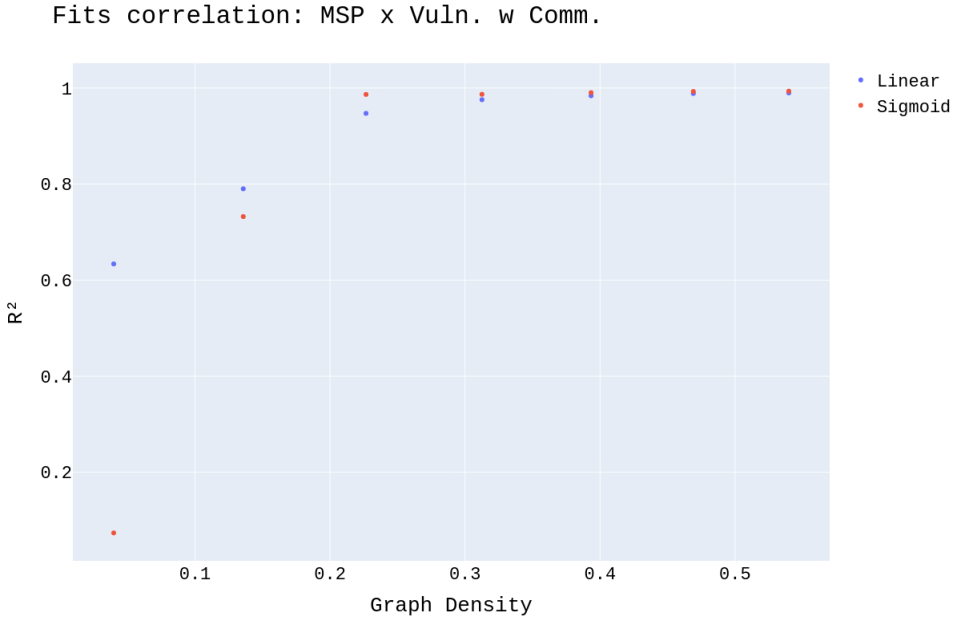
In Figure 4.23 we notice how steadily both types of fit associated R^2 grows, with the sigmoid function always with a more considerable value.

Figure 4.24 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$.



To the betweenness in Figure 4.24, we notice an oscillation of values in a short-range. We also notice how the sigmoid fit is always closer to 1 than the linear.

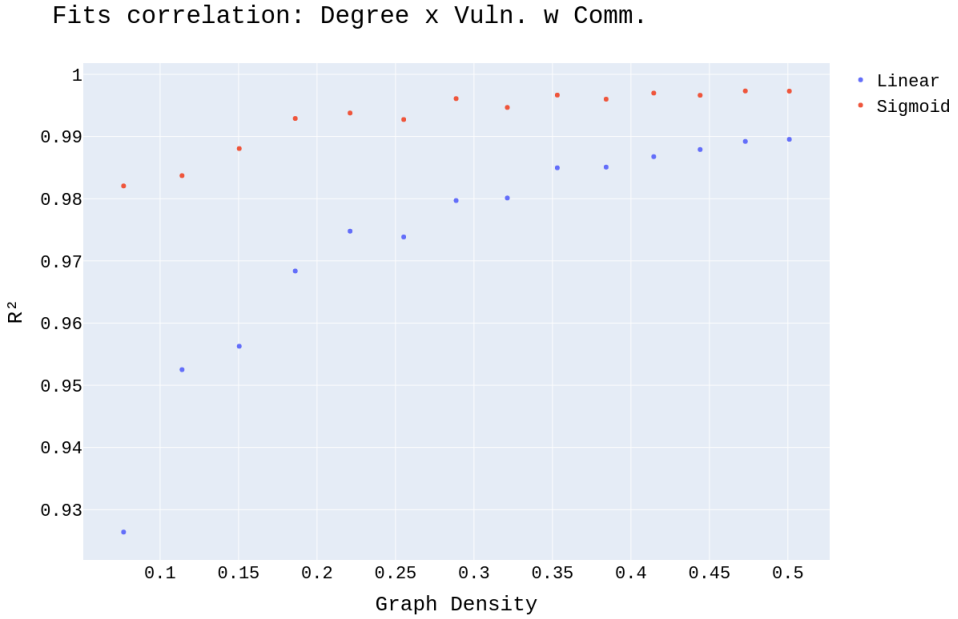
Figure 4.25 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 100$.



In Figure 4.25 we can see how the sigmoid function R^2 is always close to one, while the linear function increases with graph density from 0.5.

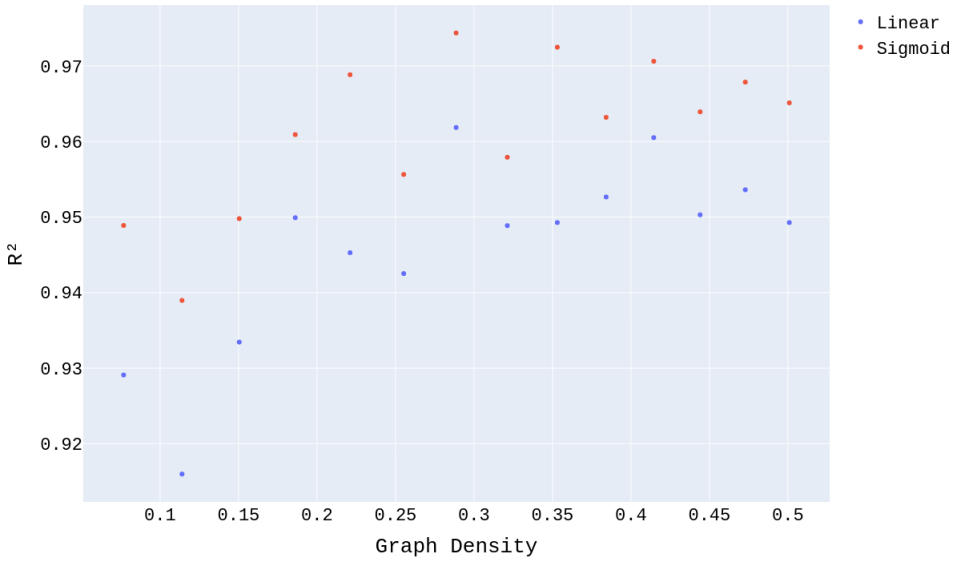
We show the results by repeating the process to generate graphs with $N = 256$.

Figure 4.26 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$.



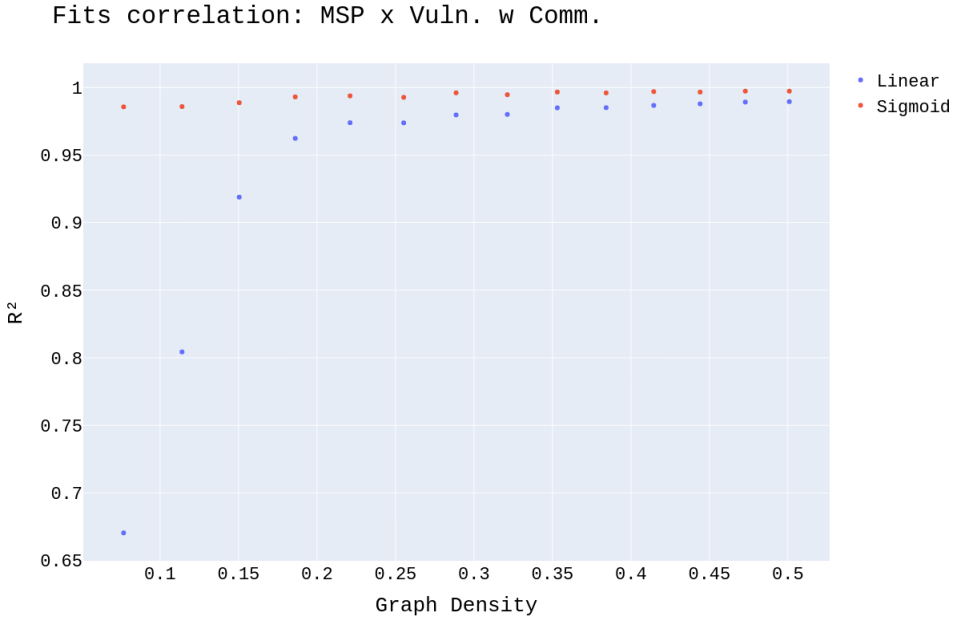
We easily notice the same steady increase in Figure 4.26 as in Figure 4.23, with sigmoid's R^2 closer to 1 than the linear function R^2.

Figure 4.27 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$.



In Figure 4.27 we can see an oscillation as well, but here we notice an increase of R^2, even with oscillations. The sigmoid function still performs better.

Figure 4.28 - Representation of R^2 vs Graph Density to randomly generated graphs with number of nodes $N = 256$.



Lastly, to the mean shortest path in Figure 4.28, we notice a predominance in the sigmoid's R^2, being always more significant than the linear function. Similar to the case with $N = 100$.

4.4 Conclusion

This work compares metrics to two graph models, representing random and scale-free networks. We compare degree, betweenness, and mean shortest path with Vulnerability with Efficiency and Vulnerability with Communicability to four configurations of generated graphs, exploring their size and densities.

This comparison resulted in distinct relations between metrics, however their behavior has something in common, in all cases the vulnerabilities increase with degree and betweenness while decreasing with the mean shortest path length. Each different graph model changes the relations between metrics, and the sharpness of those relations is closely related to the graph's density and size. We achieve better results of fits in bigger and denser graphs.

The relations obtained can be fitted to linear and sigmoid functions; the Erdős–Rényi model has a predominance of linear fits between relations, while the Barabási-Albert has a predominance of fits with the sigmoid function.

We choose what type of fit to use by looking at the R^2 associated and visualizing each scatter plot to understand the relation between metrics.

5 RANDOM WALKS: A STOCHASTIC VIEW ON THE HIERARCHY OF NODES

Since Vulnerability with Communicability takes many resources, we will mimic the diffusion-like method with a random walk on a network to get insights into the network structure, considering not only the shortest paths.

We propose an index and evaluate it using the previously presented Zachary Karate Club, Erdos-Renyi, and Barabási-Albert graphs.

5.1 Introduction

Random walks are a well-known process in network science literature (BURIONI; CASSI, 2005; YANG, 2005; COSTA; TRAVIESO, 2007). They are mathematical objects similar to the Brownian motion concept, a stochastic random process applied in interdisciplinary areas. It is interesting as a diffusion process, modeling the movement of information inside the network with a particle that can walk in any direction, independently from its previous movement.

Computational cost is a significant bottleneck when talking about advances in science. When calculating a metric, taking too many resources can be a deal-breaker for some studies that do not have access to computer clusters or supercomputers, and having an extensive network can lead us to long processing times even for the most straightforward metrics, like the shortest path.

Random Walks are stochastic processes; given a graph and a starting node, we select a neighbor of it at random and move to this neighbor; then, we select a neighbor of this node at random and move to it, repeating until we reach a certain amount of steps (BURIONI; CASSI, 2005).

This work defines an index based on random walks that tells us how many times a walk goes through each node. It is a normalized measure of frequency, which we call Passaging Index.

5.2 Methodology

The Passaging Index is a relative frequency of how many times the random walk goes through each node inside the network. Mathematically, it is defined as

$$PI_k = \frac{\text{number of passages}}{\text{total number of steps}}, \tag{5.1}$$

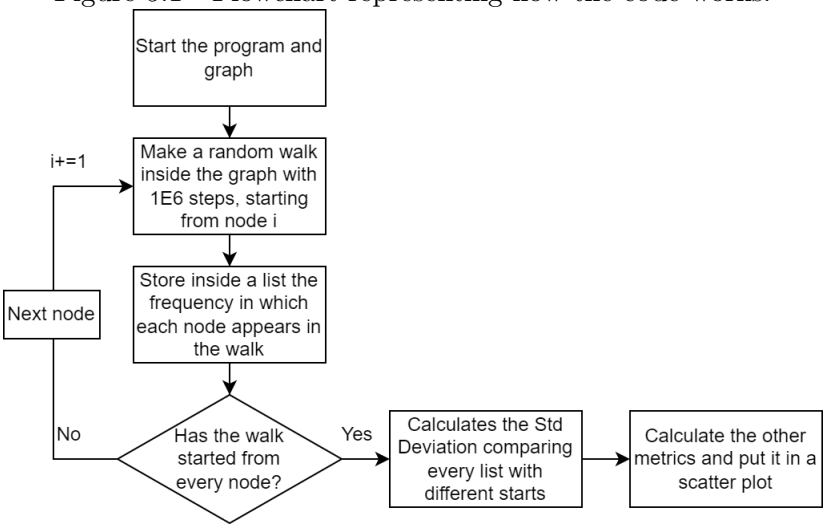
where PI_k is the Passaging Index associated with node k, the number of passages is how many times the node k shows up in the walk, and the total number of steps is the walk's length. It is important to note that the entire sum of the Passaging Index has to be one so

$$\sum_{k}^{N} PI_k = 1, \tag{5.2}$$

where k is each node, PI is the Passaging Index, and N is the set of nodes.

In the following Figure, we can see how our program operates. We initialize our network and then make a Random Walk with 10^8 steps. This Random Walk starts in the node i, where every time we complete our random walk, we start again at the node $i + 1$ until we start the Random Walk from each node.

Figure 5.1 - Flowchart representing how the code works.



Doing it this way can confirm that where we begin our walk does not make any difference; to confirm it, we calculate the standard deviation, comparing the arrays of Passaging Index.

To conclude our program, we show scatter plots to compare this index with the same

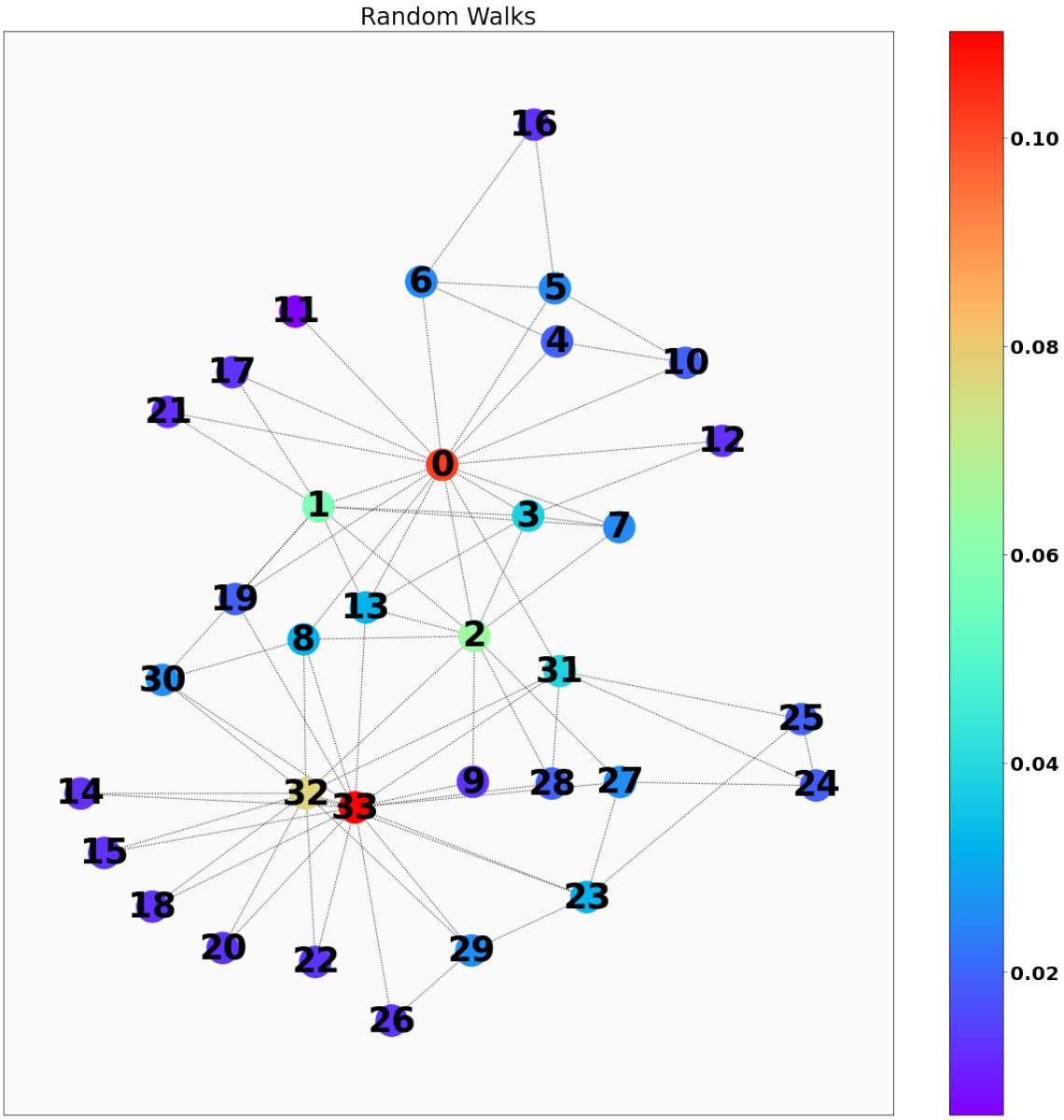
indexes from Chapter 4.

We repeat the process to randomly generated graphs created with the Erdős–Rényi and Barabási-Albert models. Those graphs are the same as our reference case in Chapter 4. We present in Appendix B the other graphs from said Chapter.

5.3 Results

We start our results by showing the color map of the Passaging Index calculated to Zachary's Karate Club graph.

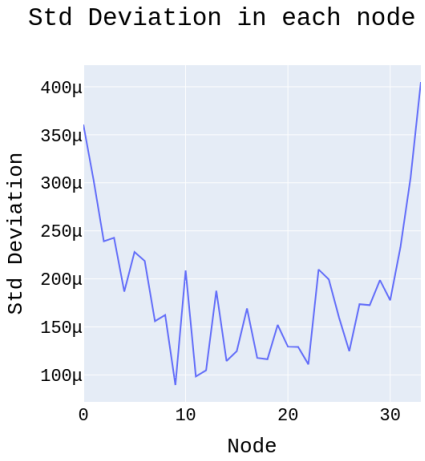
Figure 5.2 - Color map illustrating the Passaging Index to each Zachary Karate Club network node.



This color map is very similar to the Vulnerability with Communicability color map. Both maps highlight the nodes 0 and 33; however, the results are very different when comparing other nodes.

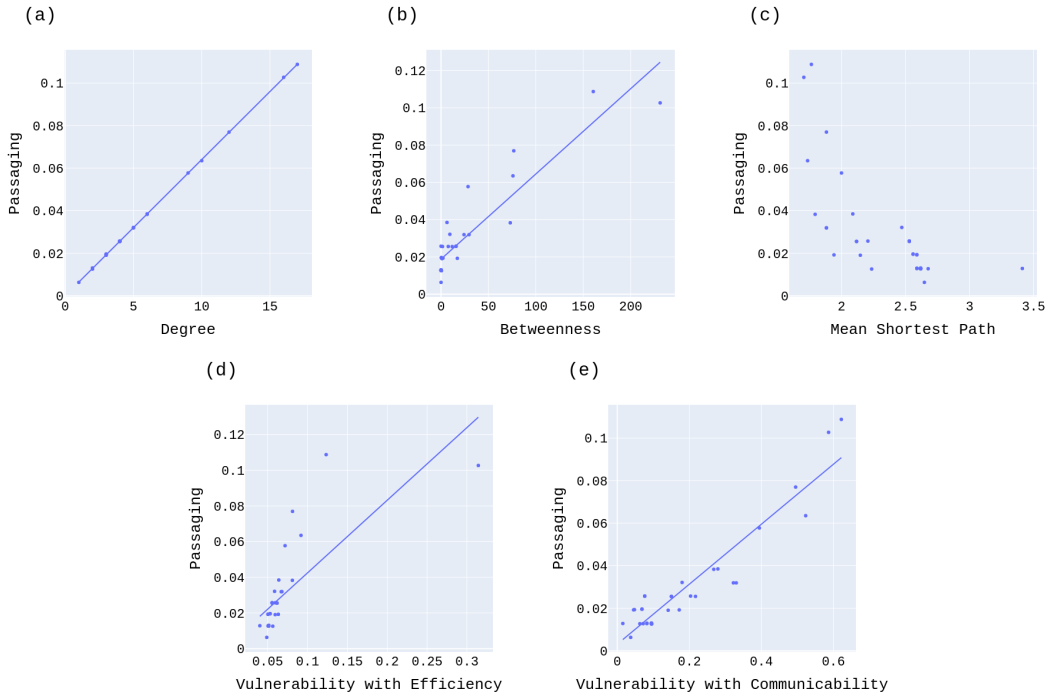
The previous Figure was made with a start point at the node 0, and to make sure that the starting point is not relevant to the final result, we make a long walk with 10^8 steps. Nonetheless, we calculate the standard deviation comparing each starting point to quantify this idea. The Figure 5.3 illustrate this.

Figure 5.3 - Standard Deviation calculated to compare the arrays made with the Passaging Index of each node. Where μ represents 10^{-6}. Made with the Zachary's Karate Club.



Here we can see how low the values of each standard deviation are, compared to the Passaging Index of each node. The mean value of the Passaging Index is 0.02, while the mean value of the standard deviation is 0.0002. So, the standard deviation is small when compared to the index itself.

Figure 5.4 - Zachary's Karate Club Passaging Index relations.

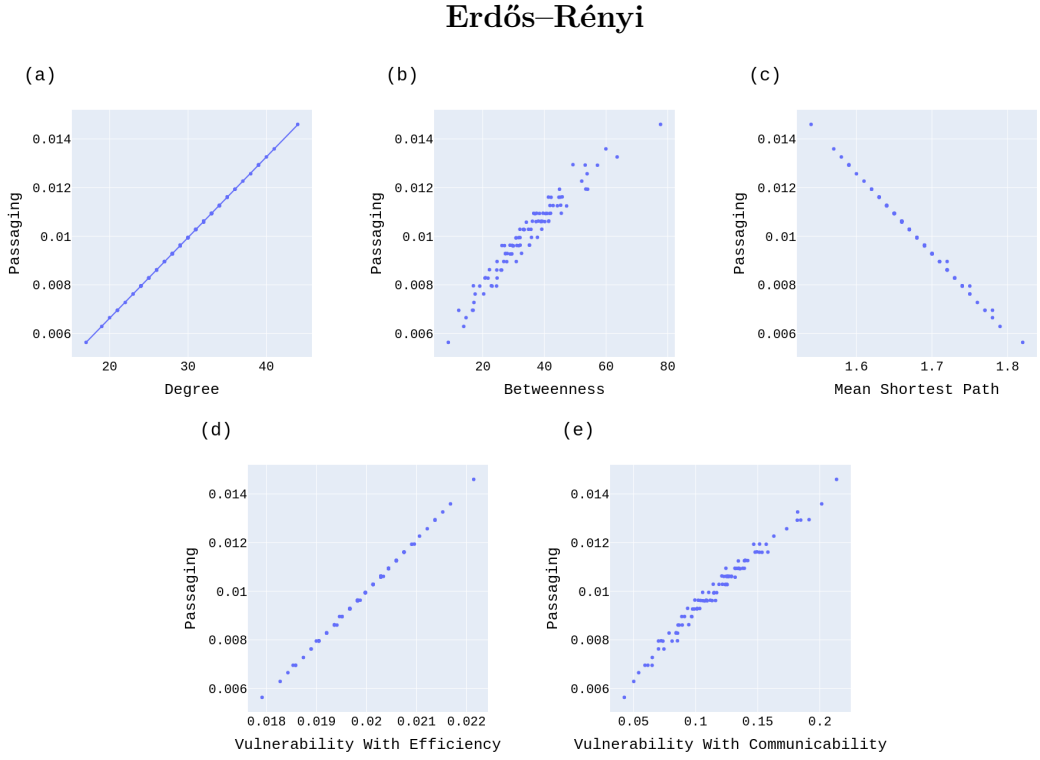


In Figure 5.4, (a) relates the Passaging Index with the degree, where we achieved a correlation coefficient equal to 1. (b) relates the Passaging Index with the Betweenness; the image shows a linear regression with a correlation coefficient equal 0.834. (c) relates the Passaging Index with the Mean Shortest Path; there is no fit since their relationship is unclear. (d) is the Passing Index with the Vulnerability with Efficiency, with a linear regression resulting in a correlation coefficient of 0.574. (e) is the Passing Index with the Vulnerability with Communicability, with a linear regression resulting in a correlation coefficient of 0.889.

Those results indicate that the Passaging Index is closely related to the degree with strong linearity. So, we further investigate different types of graphs.

Beginning with the Erdős–Rényi model, we borrow the graphs from 4 to compare with the Passaging Index.

Figure 5.5 - Scatter plots to graph with following characteristics: N = 100, L = 1475 Density = 0.3, N = 100, L = 1547, $< c > = 0.3$, $< k > = 29.50$, D = 2, $< l > = 1.68$.

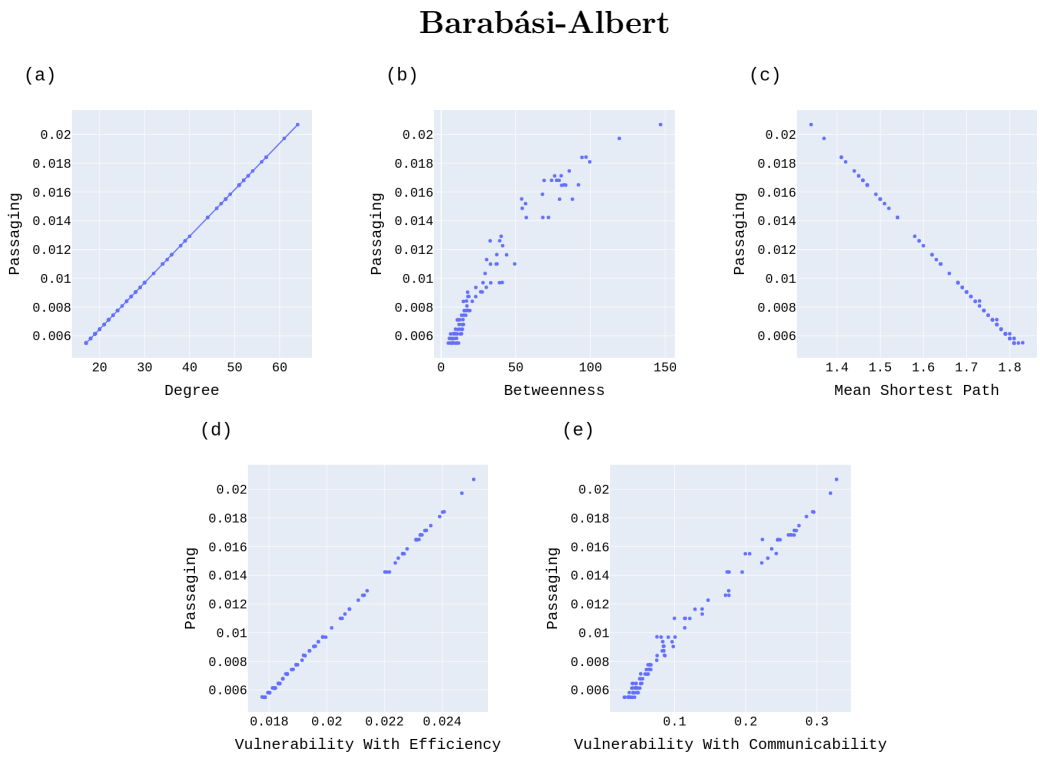


In Figure 5.5 we can see how the linear relationship of the Passaging Index is not

only valid to the degree, extending itself to all other metrics in the Erdős–Rényi model. We also calculate those metrics and relationships to the other graphs in Chapter 4. However, the results are similar, so we present them in Appendix B.

As expected, the Passaging Index increases with degree, betweenness, and both vulnerabilities, while decreasing with the mean shortest path.

Figure 5.6 - Scatter plots to graph with following characteristics: N = 100, L = 1547 Density = 0.3, N = 100, L = 1547, $< c > = 0.42$, $< k > = 30.94$, D = 3, $< l > = 1.67$.



We can see in Figure 5.6 that the Passaging Index increases with degree, betweenness, and both vulnerabilities, while decreasing with the mean shortest path.

Just like in Chapter 4, the relationship between metrics changes to the Barabási-Albert model. Not all of them have a solid linear representation and change even more to other densities, as shown in Appendix B.

5.4 Conclusion

We defined a stochastic metric called Passaging Index, that considers the frequency in which a long random walk goes through each node.

Analyzing the Zachary's karate club graph, we found that the relation between degree and Passaging Index is perfectly linear. Since it is a small network, we investigate it further by generating random graphs with two models: random and scale-free.

Each model had the same linear relation between degree and Passaging Index. However, when analyzing the other relations, the scale-free network had more complex interactions, with scattered points over the plot.

Since the Passaging Index incorporates random walks without any memory, we are instigated to investigate other types of random walks, like Self-Avoiding Walks and other similar walks.

6 FINAL REMARKS

Vulnerability is a key word in Disaster Science (UNDRR, 2022). Vulnerability is also a concept, and a tool in Network Science (LATORA; MARCHIORI, 2004; GOLDSHTEIN et al., 2004). In this work, we analyzed the vulnerabilities of networks based on a multidisciplinary approach, considering not only the shortest paths between the pairs of network nodes.

First, we defined a new metric of network vulnerability - applying Communicability as performance instead of Efficiency. Then, we show the broad difference between them by highlighting different nodes as vulnerable: Efficiency uses only the shortest paths, and Communicability uses all paths that go through a node.

We can see that Vulnerability with Communicability has a broader range than Vulnerability with Efficiency. It is also important to note that the Vulnerability with Communicability has more nodes with a vulnerability closer to the maximum value.

However, our first analysis of network vulnerabilities only uses a small network. To take better conclusions, we explore bigger graphs, randomly generated using two graph models, the Random Network (Erdős–Rényi model) and the Scale-free Network (Barabási-Albert model).

For each model, we generate four configurations of random graphs, exploring how the increase of nodes and decrease of edges can impact our results. We keep the number of nodes equal when comparing models and an approximate number of edges between each iteration.

Then, we compare each Vulnerability with well-known metrics in the literature as the degree, betweenness, and mean shortest path.

The comparison is made in a scatter plot, putting each Vulnerability against the degree, betweenness, and mean shortest path. This resulted in distinct relations between metrics; however, their behavior has something in common: In all cases, the vulnerabilities increase with degree and betweenness while decreasing with the mean shortest path length.

Each different graph model changes the relations between metrics, and the sharpness of those relations is closely related to the graph's density and size. We achieve better results of fits in bigger and denser graphs.

The relations obtained can be fitted to linear and sigmoid functions; the Erdős–Rényi model has a predominance of linear relations, while the Barabási-Albert has a predominance of fits with the sigmoid function.

In a few words, our recommendation is: When longer paths are relevant to the analysis, use the Vulnerability with Communicability, and when only the shortest path is relevant, use the Vulnerability with Efficiency.

In the last chapter, we define the Passaging Index. The Passaging Index is a stochastic index that considers the frequency of a long random walk that goes through each node.

We analyze Zachary's karate club graph and the same graphs presented in Chapter 4. We conclude that our Passaging Index has a perfect linear correlation with the degree, reaching a $R^2 = 1$ to all graphs tested.

Besides that, we can see the differences arising in each graphs model. Each model had its characteristics when comparing the Passaging Index with other metrics such as Betweenness, Mean Shortest Path, and Vulnerabilities.

The definition of the Passaging Index opens the way to investigate walks with memory, like Self-Avoiding Walks, since our study only used a random walk without any memory.

We intend to explore these other types of random walks in future works. Also, we aim to study innovative ways of exploring those network metrics using graphs and High-performance computing.

REFERENCES

ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. **Reviews of Modern Physics**, v. 74, n. 1, p. 47, 2002. 1, 10

BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **Science**, v. 286, n. 5439, p. 509–512, 1999. 9

BARABÁSI, A. L.; BONABEAU, E. Scale-free networks. **Scientific American**, v. 288, n. 5, p. 60–69, 2003. ISSN 00368733. 1

BARABáSI, A.-L. **Linked-the new science of networks**. [S.l.]: Perseus Books Group, 2002. ISBN 0738206679,9780738206677. 6

BENDER, S. G. W. E. A. **Lists, decisions and graphs**. [s.n.], 2015. (Lectures in Discrete Mathematics 2). Available from: <libgen.li/file.php?md5=a6415ebbae9596af1d7fe8656fc9bc82>. 12

BURIONI, R.; CASSI, D. Random walks on graphs: ideas, techniques and results. **Journal of Physics A: Mathematical and General**, v. 38, n. 8, p. R45, 2005. 55

CARLSON, S. C. **Konigsberg bridge problem**. 2021. Available from: <https://www.britannica.com/science/Konigsberg-bridge-problem>. Access in: 15 Mar. 2021. 5

CHEN, G.; DONG, Z. Y.; HILL, D. J.; ZHANG, G. H.; HUA, K. Q. Attack structural vulnerability of power grids: a hybrid approach based on complex networks. **Physica A: Statistical Mechanics and its Applications**, v. 389, n. 3, p. 595–603, 2010. 2, 15

CHEN, H.; ZHANG, L.; RAN, L. Vulnerability modeling and assessment in urban transit systems considering disaster chains: a weighted complex network approach. **International Journal of Disaster Risk Reduction**, v. 54, p. 102033, 2021. ISSN 2212-4209. Available from: <https://www.sciencedirect.com/science/article/pii/S2212420920315351>. 13

COSTA, L. d. F.; RODRIGUES, F. A.; TRAVIESO, G.; BOAS, P. R. V. Characterization of complex networks: a survey of measurements. **Advances in Physics**, v. 56, n. 1, p. 167–242, 2007. 25

COSTA, L. da F.; TRAVIESO, G. Exploring complex networks through random walks. **Physical Review E**, APS, v. 75, n. 1, p. 016102, 2007. 14, 55

ERDŐS, P.; RÉNYI, A. On random graphs. **Publicationes Mathematicae**, v. 6, p. 290–297, 1959. 1, 9

ERDőS, P.; RéNYI, A. On the evolution of random graphs. **Publications of the Mathematical Institute of the Hungarian Academy of Sciences**, v. 5, n. 1, p. 17–60, 1960. 1, 9

ESTRADA, E.; HATANO, N. Communicability in complex networks. **Physical Review E - Statistical, Nonlinear, and Soft Matter Physics**, v. 77, n. 3, p. 1–12, 2008. ISSN 15393755. 2, 13, 16, 17, 25

EULER, L. Solutio problematis ad geometriam situspertinentis. 1736. Available from: <https://scholarlycommons.pacific.edu/cgi/viewcontent.cgi?article=1052&context=euler-works>. 1

G1. **Apagão que gerou crise energética no Amapá deverá ser investigado e julgado só na esfera federal**. 2020. Available from: <t.ly/IoGx>. Access in: 19 Mar. 2021. 1

GALBRUN, E.; PELECHRINIS, K.; TERZI, E. Urban navigation beyond shortest route: The case of safe paths. **Information Systems**, v. 57, p. 160–171, 2016. ISSN 0306-4379. Available from: <https://www.sciencedirect.com/science/article/pii/S0306437915001854>. 2

GOLDSHTEIN, V.; KOGANOV, G. A.; SURDUTOVICH, G. I. Vulnerability and hierarchy of complex networks. p. 1–4, 2004. Available from: <http://arxiv.org/abs/cond-mat/0409298>. 2, 13, 15, 18, 19, 25, 63

GUARDIAN, T. **Why the cold weather caused huge Texas blackouts – a visual explainer**. Guardian News and Media, Feb 2021. Available from: <https://www.theguardian.com/us-news/2021/feb/20/texas-power-grid-explainer-winter-weather>. Access in: 19 Mar. 2021. 1

HOLME, P.; KIM, B. J.; YOON, C. N.; HAN, S. K. Attack vulnerability of complex networks. **Physical Review E**, v. 65, n. 5, p. 056109, 2002. 2, 15

HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the National Academy of Sciences**, v. 79, n. 8, p. 2554–2558, 1982. 1, 15

JEONG, H.; TOMBOR, B.; ALBERT, R.; OLTVAI, Z. N.; BARABÁSI, A.-L. The large-scale organization of metabolic networks. **Nature**, v. 407, n. 6804, p. 651–654, 2000. 1, 15

LATORA, V.; MARCHIORI, M. Efficient behavior of small-world networks. **Physical Review Letters**, v. 87, n. 19, p. 198701–1–198701–4, 2001. ISSN 10797114. 13, 16, 18, 19, 25

_____. Vulnerability and protection of critical infrastructures. p. 11–14, 2004. Available from: <http://arxiv.org/abs/cond-mat/0407491{%}0Ahttp://dx.doi.org/10.1103/PhysRevE.71.015103>. 2, 13, 15, 17, 25, 63

LIMA, A.; STANOJEVIC, R.; PAPAGIANNAKI, D.; RODRIGUEZ, P.; GONZáLEZ, M. C. Understanding individual human mobility patterns. **Nature**, v. 453, n. 7196, p. 779–782, 2008. ISSN 14764687. 2, 15, 16

MISHKOVSKI, I.; BIEY, M.; KOCAREV, L. Vulnerability of complex networks. **Communications in Nonlinear Science and Numerical Simulation**, v. 16, n. 1, p. 341–349, 2011. 2, 13, 15

ROCCO, C. M.; RAMIREZ-MARQUEZ, J. E. Vulnerability metrics and analysis for communities in complex networks. **Reliability Engineering &amp System Safety**, Elsevier BV, v. 96, n. 10, p. 1360–1366, oct. 2011. Available from: <https://doi.org/10.1016/j.ress.2011.03.001>. 2, 15

SANTOS, L. B.; GARG, T.; JORGE, A. A.; LONDE, L. R.; REANI, R. T.; BACELAR, R. B.; SOKOLOV, I. M. Vulnerability analysis in complex networks under a flood risk reduction point of view. **arXiv**, p. 1–9, 2020. 15

SANTOS, L. B. et al. Desastres naturais de origem hidrológica e impactos no setor de transportes-o caso de março de 2015 em São José dos Campos-SP. In: BRAZILIAN SYMPOSIUM ON WATER RESOURCES 2015. **Proceedings...** Brasilia, 2015. 1, 15

SOARES, G.; SANTOS, L. Beyond the shortest path - a topological vulnerability analysis. In: . [S.l.: s.n.], 2021. p. 1–9. 15

TSONIS, A. A.; SWANSON, K. L.; ROEBBER, P. J. What do networks have to do with climate? **Bulletin of the American Meteorological Society**, v. 87, n. 5, p. 585–595, 2006. ISSN 00030007. 1, 15

UNITED NATIONS FOR DISASTER RISK REDUCTION (UNDRR). **Disaster**. 2022. Available from: <https://www.undrr.org/terminology>. Access in: 21 Mar. 2022. 1, 63

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of 'small-world'networks. **Nature**, v. 393, n. 6684, p. 440–442, 1998. 1

WIKIPEDIA. **Seven Bridges of Königsberg**. 2021. Available from: <https://en.wikipedia.org/w/index.php?title=Seven_Bridges_of_K%C3%B6nigsberg&oldid=1059621186>. Access in: 24 Jan. 2022. 6

YANG, S.-J. Exploring complex networks by walking on them. **Physical Review E**, v. 71, n. 1, p. 016107, 2005. 14, 55

ZACHARY, W. W. An information flow model for conflict and fission in small groups. **Journal of Anthropological Research**, v. 33, n. 4, p. 452–473, 1977. 19, 20

APPENDIX A - ADDITIONAL FIGURES FROM CHAPTER 4

We use this appendix to show different cases from Chapter 4.

A.1 Erdős–Rényi

A.1.1 N = 100, L = 584

Figure A.1 - Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4.

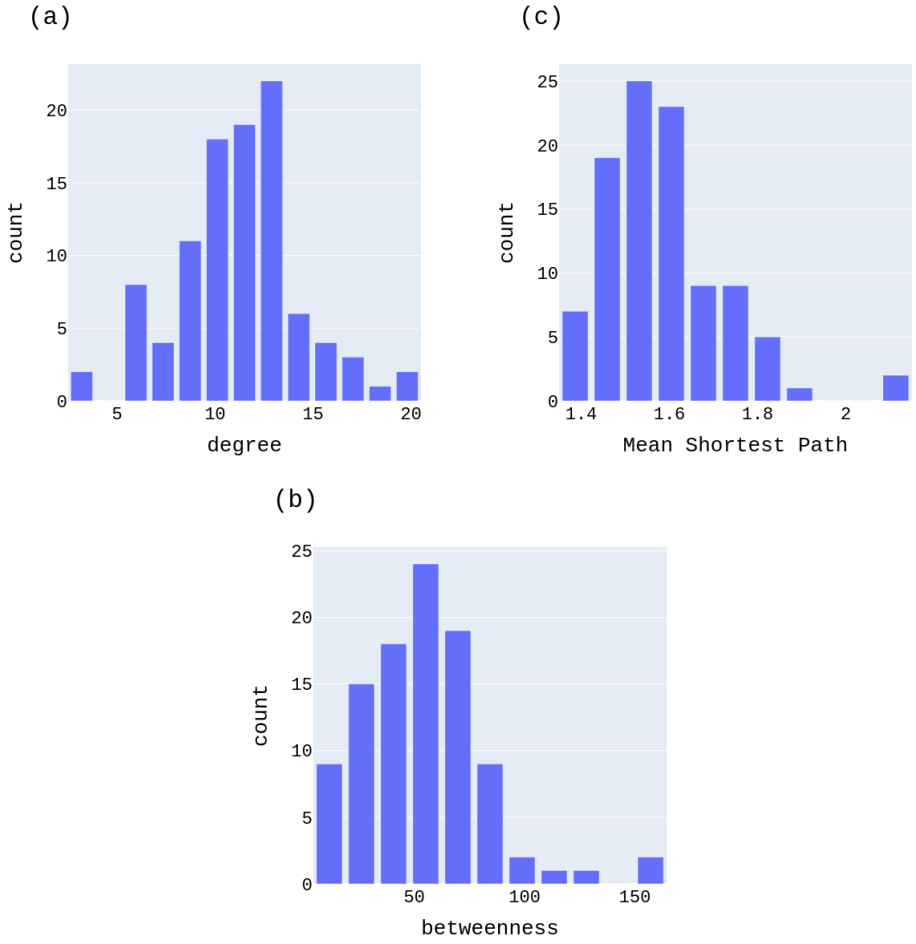
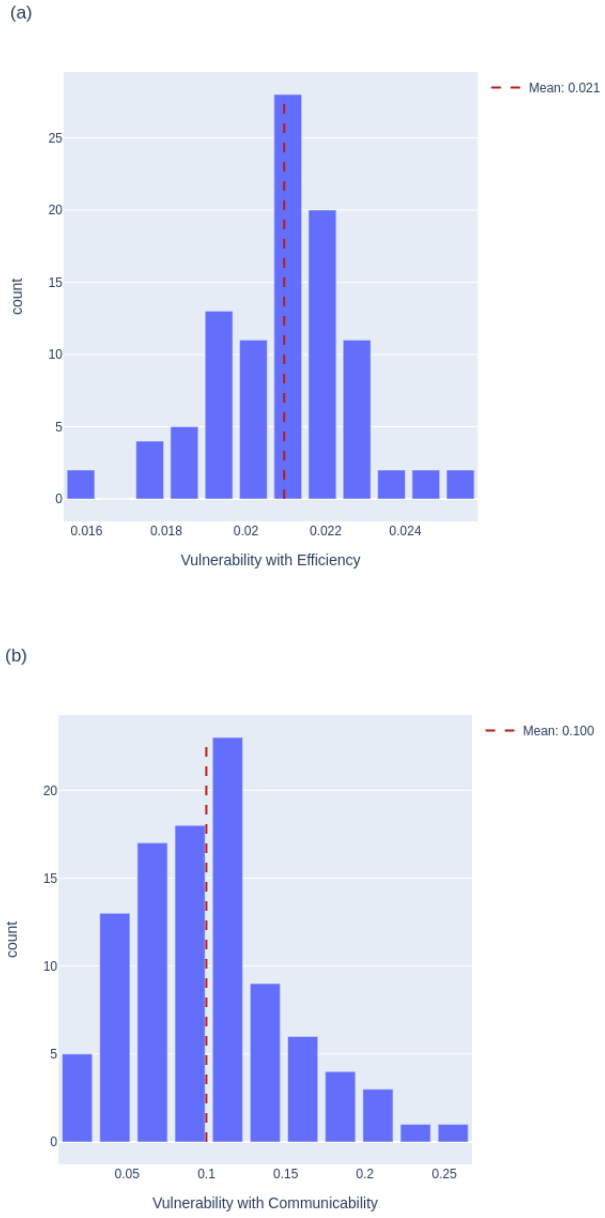


Figure A.2 - Erdős–Rényi distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value.



A.1.2 N = 256, L = 1329

Figure A.3 - Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4.

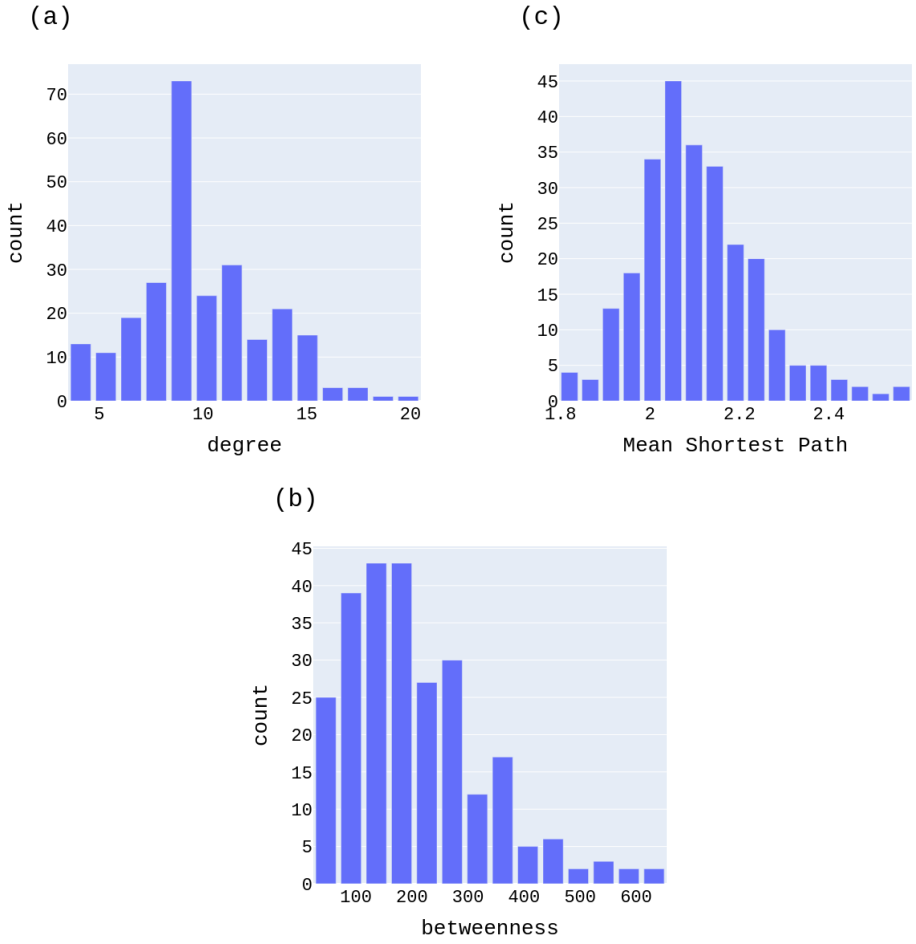
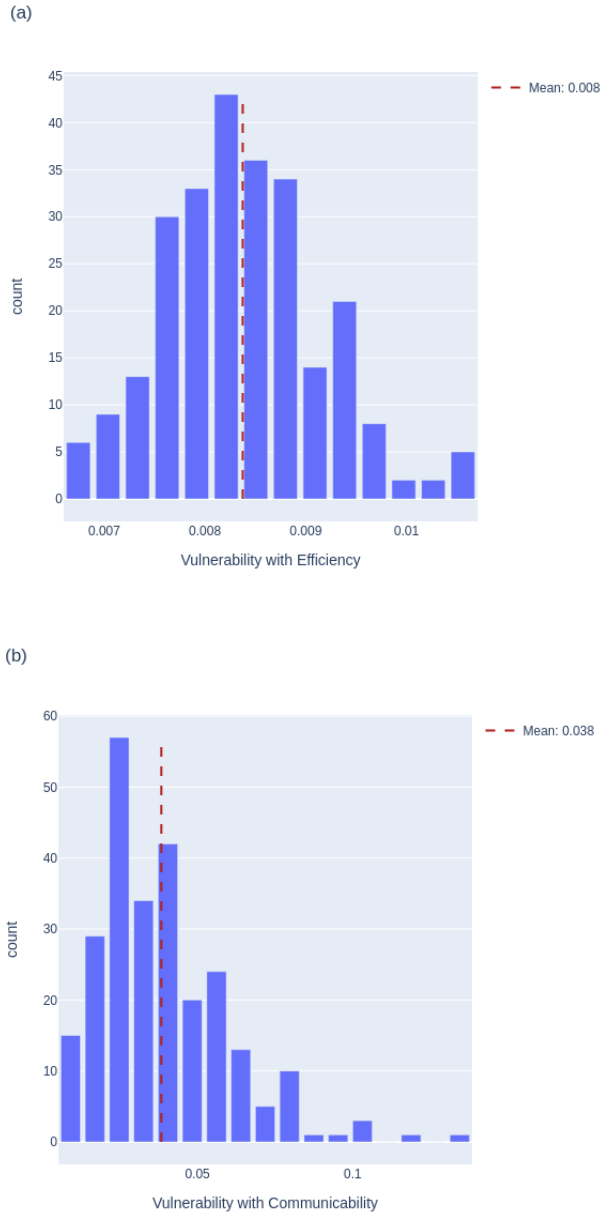


Figure A.4 - Erdős–Rényi distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value.



A.1.3 N = 2000, L = 99685

Figure A.5 - Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4.

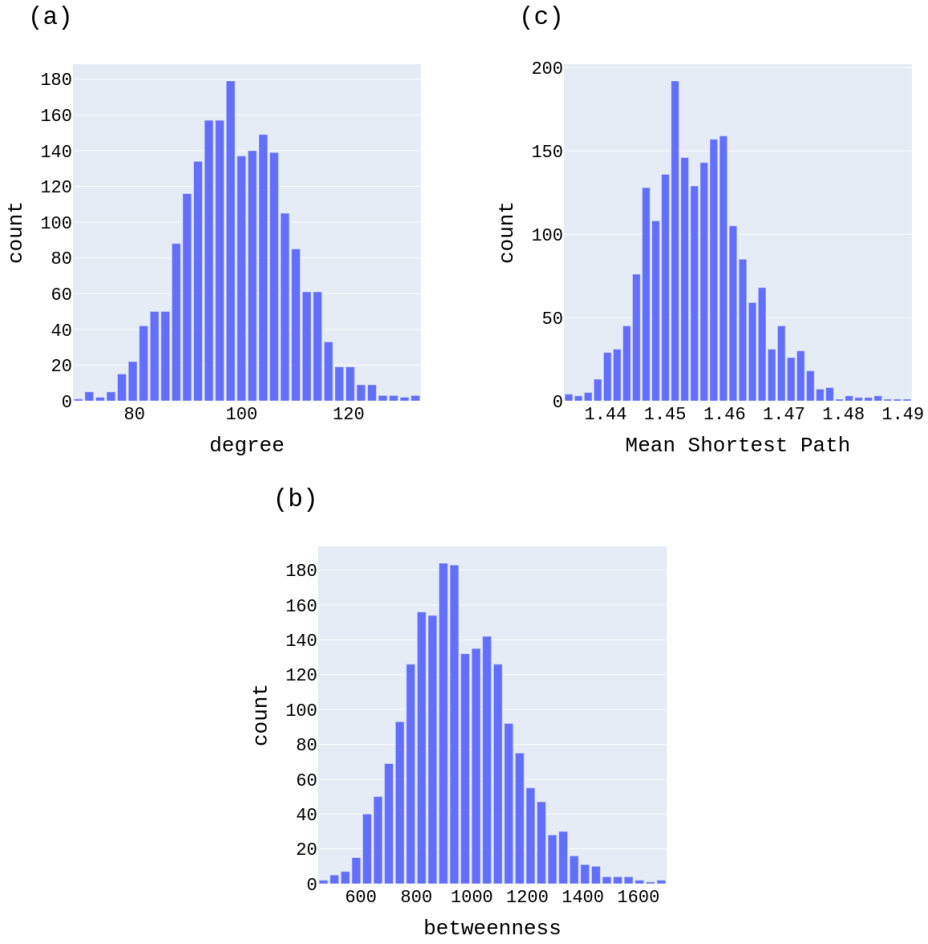
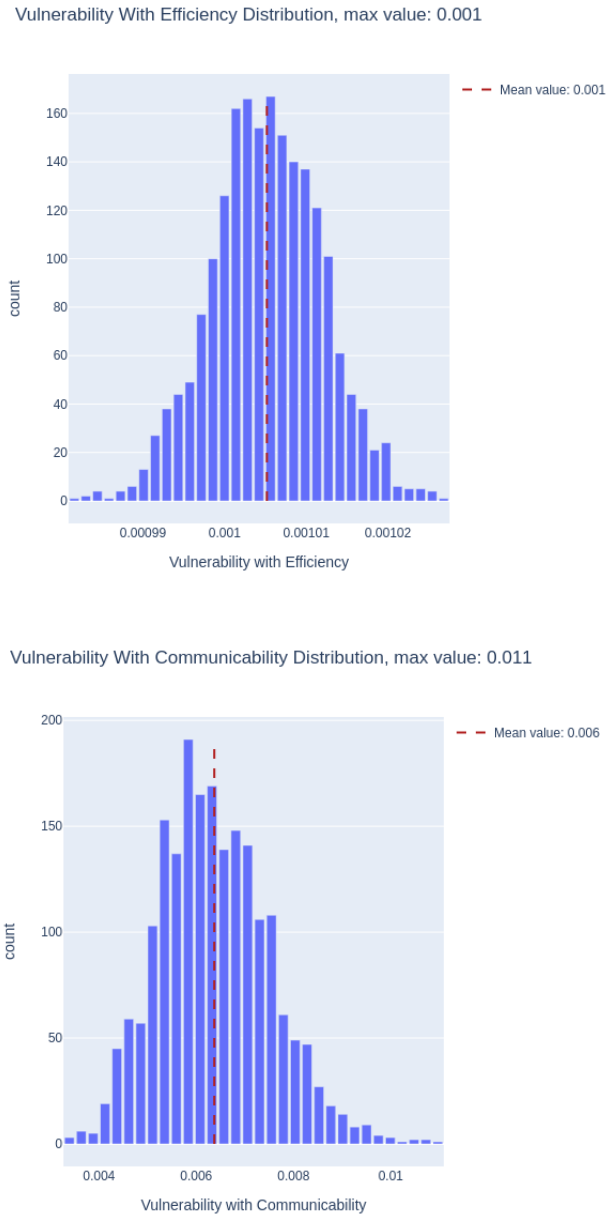


Figure A.6 - Erdős–Rényi distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value.



A.2 Barabási-Albert

A.2.1 N = 100, L = 584

Figure A.7 - Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4.

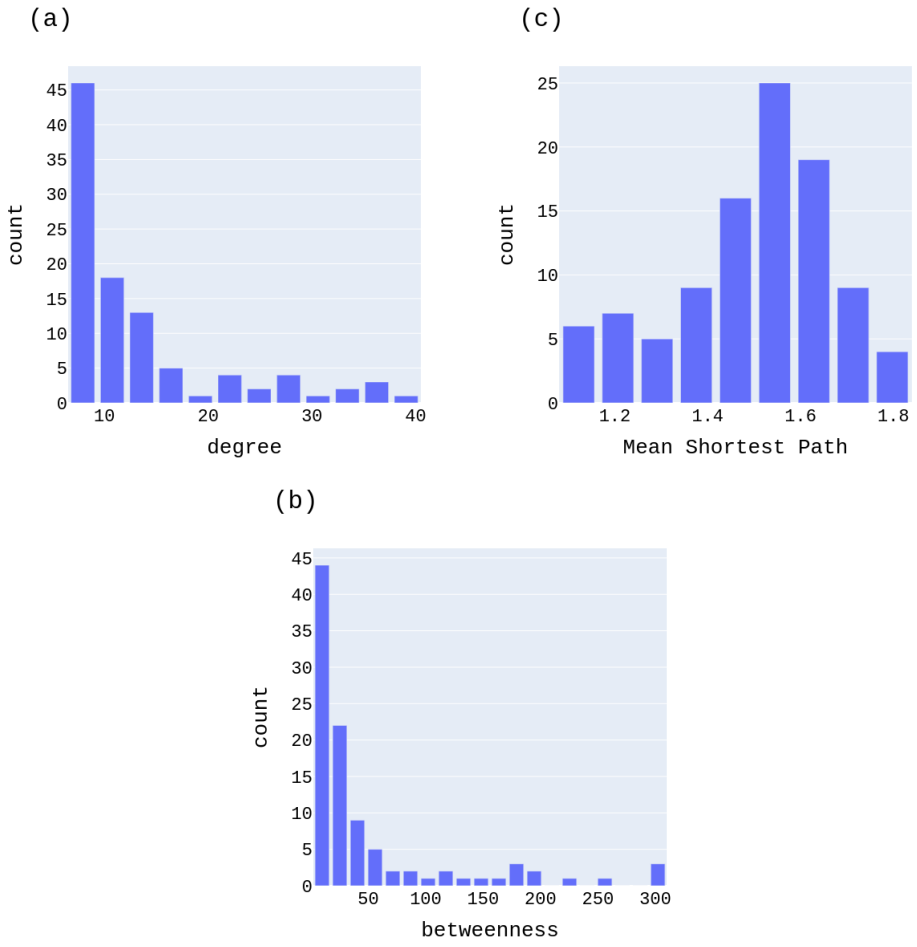
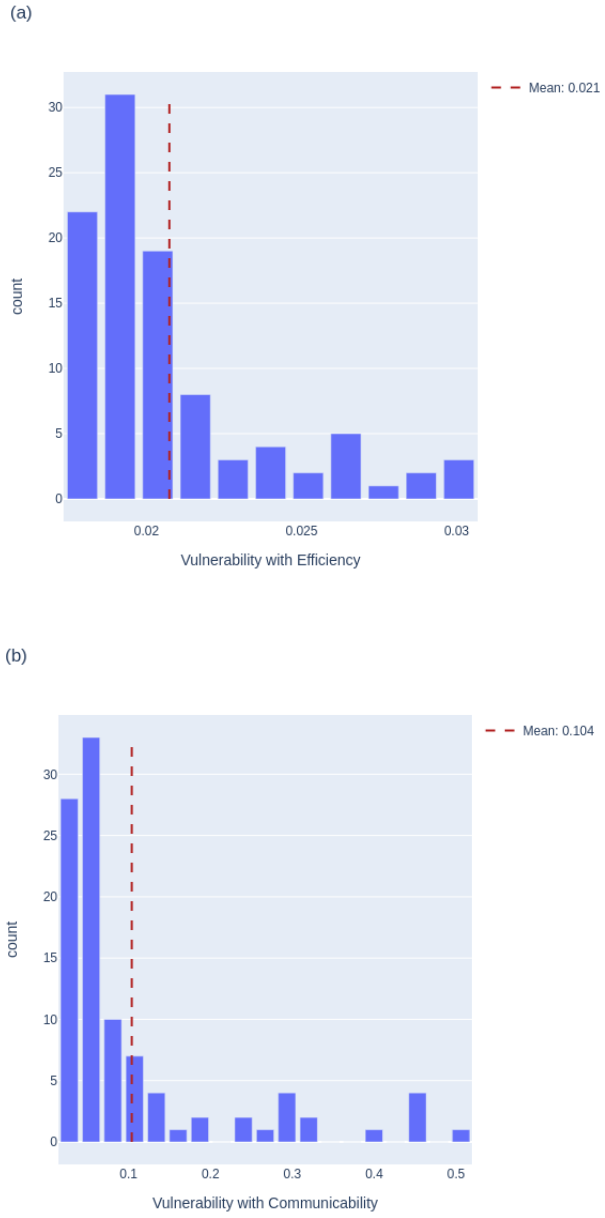


Figure A.8 - Barabási-Albert distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value.



A.2.2 N = 256, L = 1329

Figure A.9 - Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4.

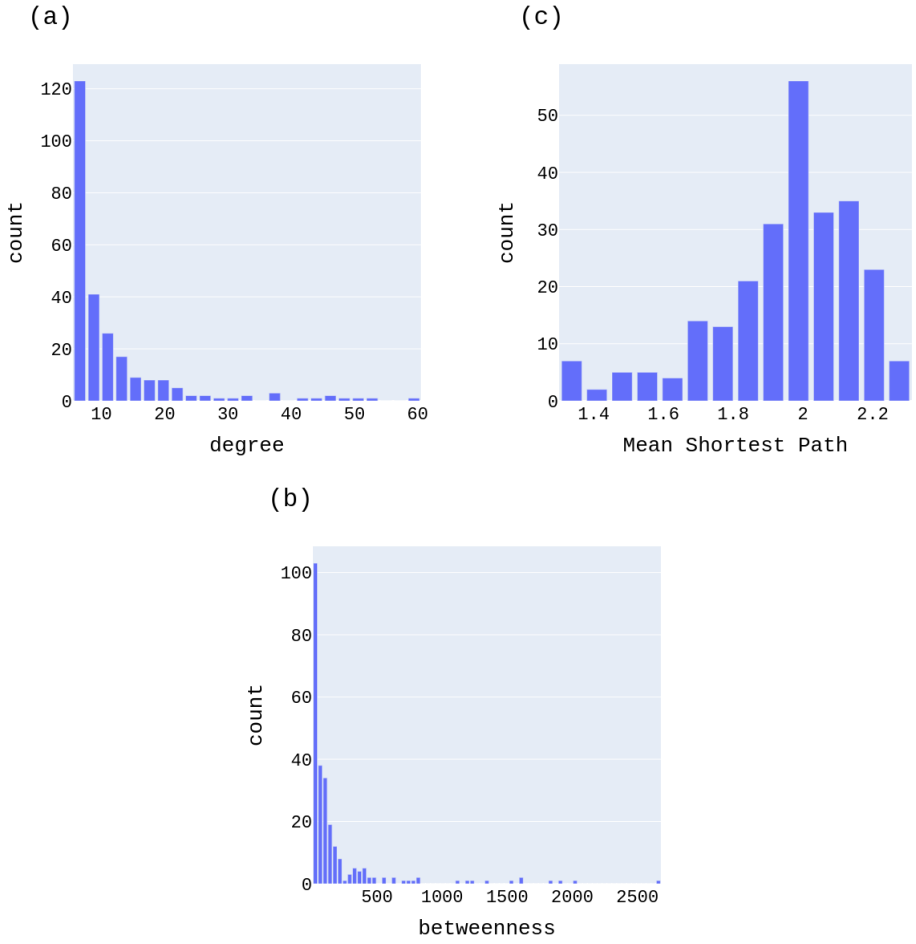
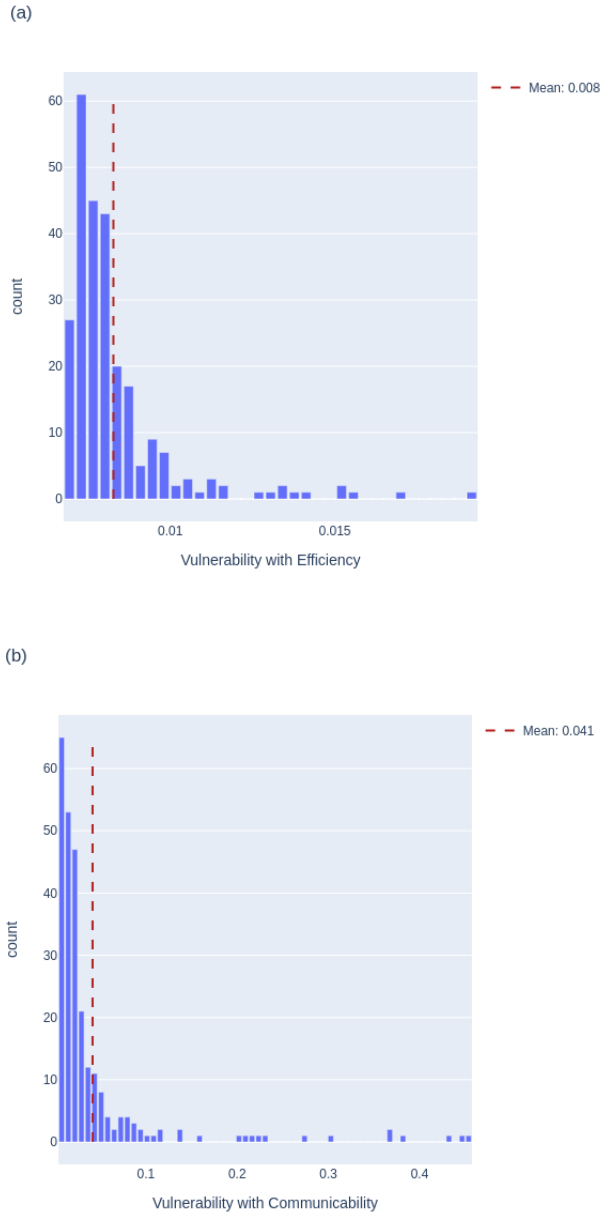


Figure A.10 - Barabási-Albert distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value.



A.2.3 N = 2000, L = 99685

Figure A.11 - Histogram of the degree, betweenness and mean shortest path distributions to each model presented before. We calculate those distributions to the same graphs in Chapter 4.

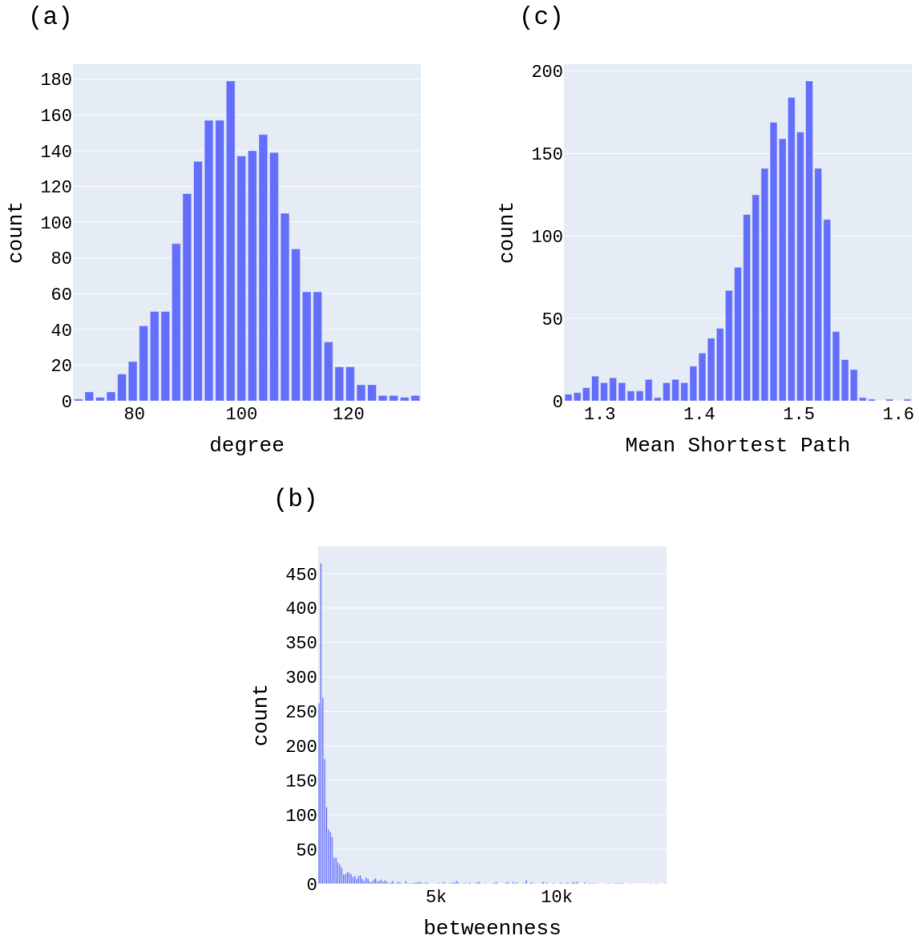
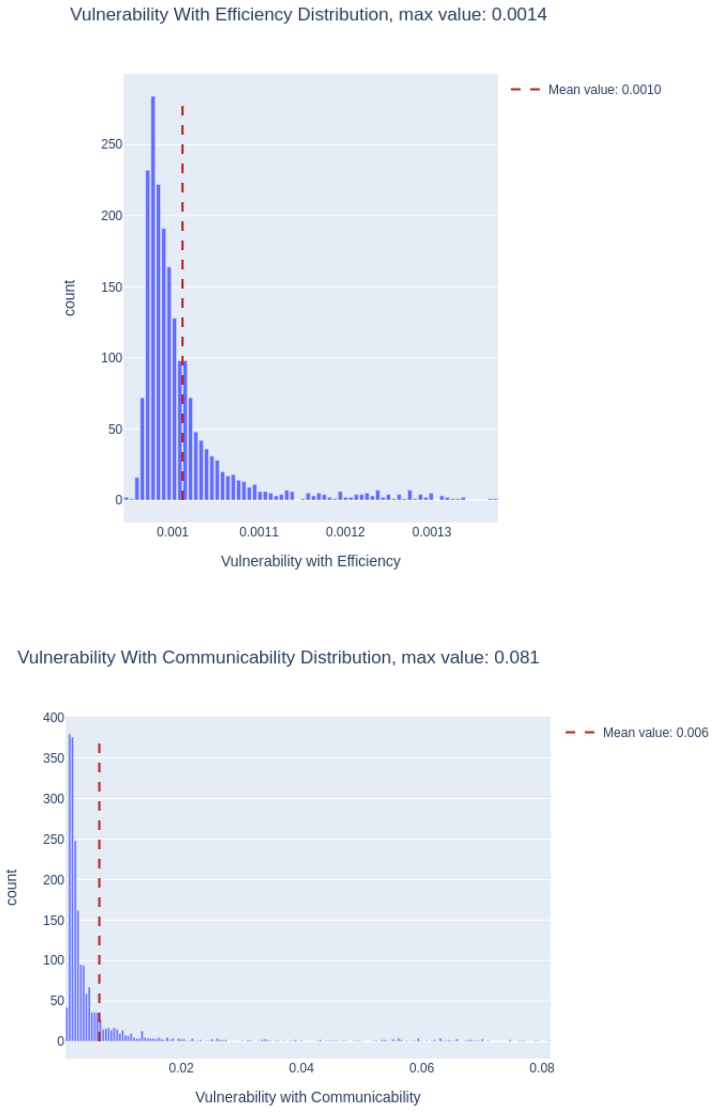


Figure A.12 - Barabási-Albert distribution of each vulnerabilities shows how many times a value inside the bin's interval appears. The line represents the sample's mean value.



APPENDIX B - ADDITIONAL FIGURES FROM CHAPTER 5

We use this appendix to show different cases from Chapter 5.

B.1 Erdős–Rényi

Figure B.1 - Scatter plots to graph with following characteristics: N = 100, L = 584, Density = 0.11, $< c > = 0.11$, $< k > = 11.68$, D = 4, $< l > = 2.08$.

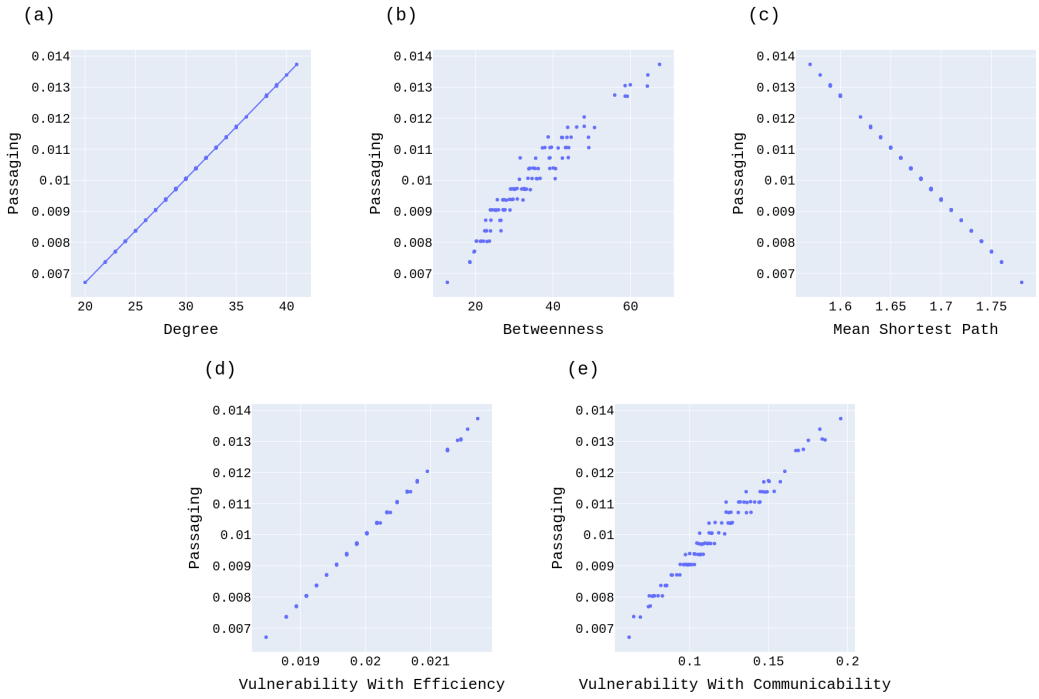


Figure B.2 - Scatter plots to graph with following characteristics: N = 256, L = 1329, Density = 0.04, $< c > = 0.04$, $< k > = 10.38$, D = 4, $< l > = 2.61$.

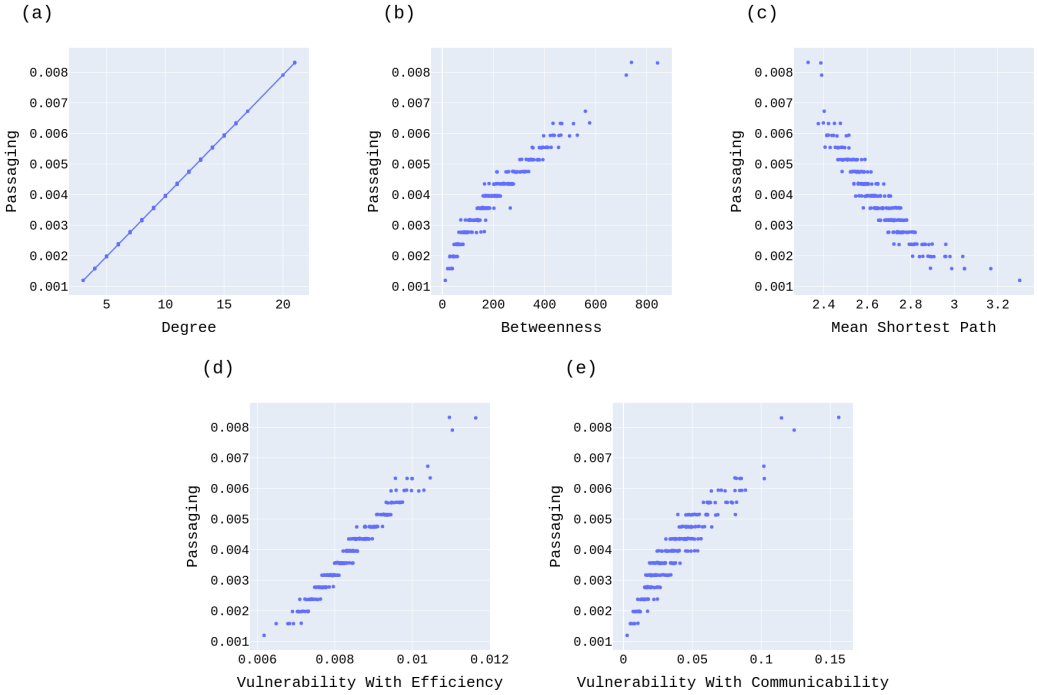
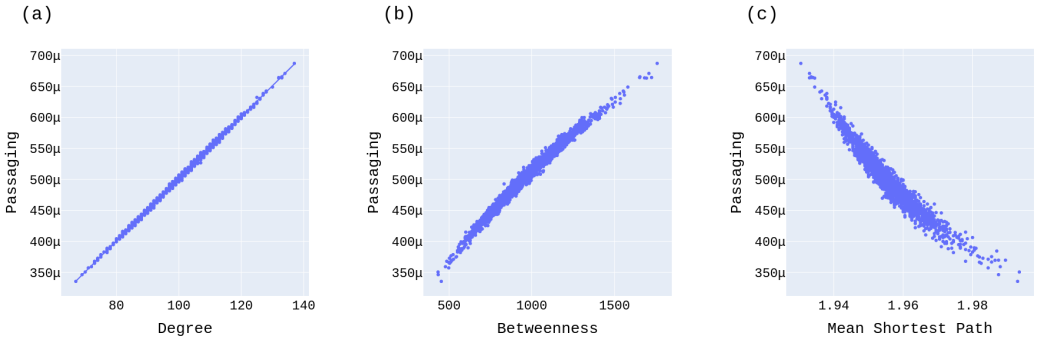


Figure B.3 - Scatter plots to graph with following characteristics: N = 2000, L = 99685, Density = 0.05, $< c > = 0.05$, $< k > = 99.68$, D = 3, $< l > = 1.95$.



B.2 Barabási-Albert

Figure B.4 - Scatter plots to graph with following characteristics: N = 100, L = 672, Density = 0.13, $< c > = 0.23$, $< k > = 13.44$, D = 3, $< l > = 1.99$.

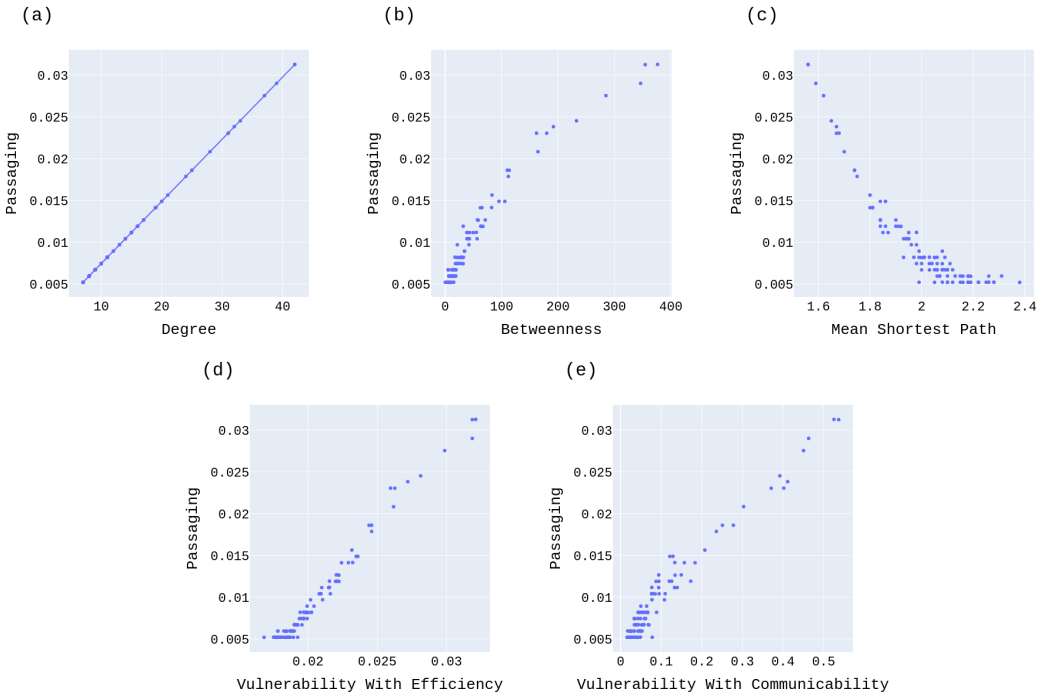


Figure B.5 - Scatter plots to graph with following characteristics: N = 256, L = 1515, Density = 0.05, $< c > = 0.10$, $< k > = 11.83$, D = 4, $< l > = 2.45$.

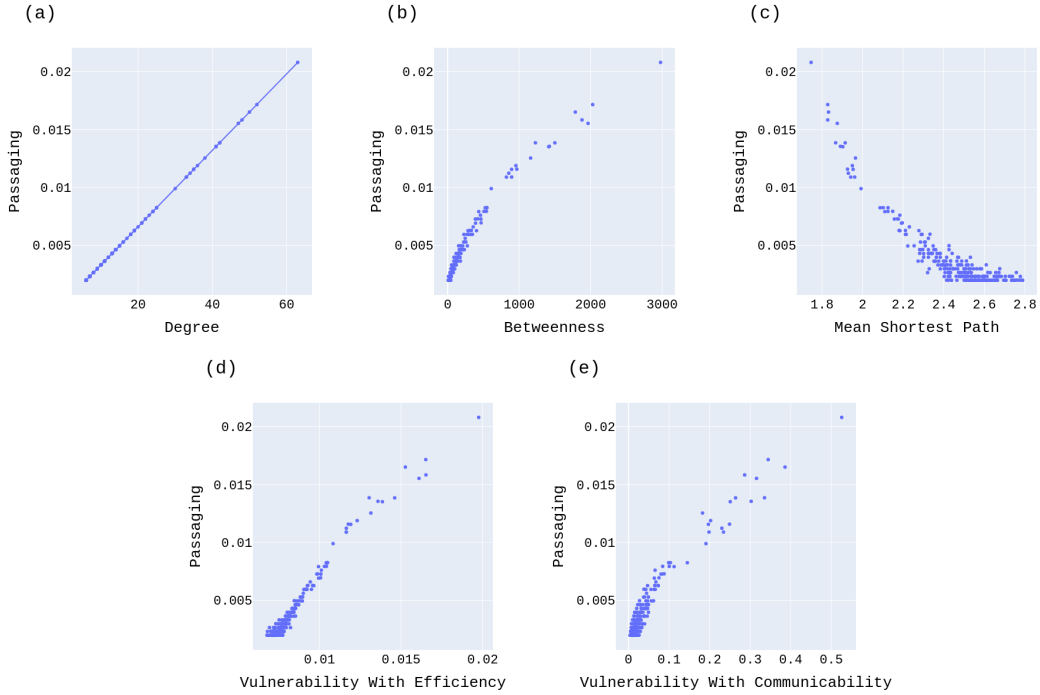


Figure B.6 - Scatter plots to graph with following characteristics: N = 2000, L = 98725, Density = 0.05, $< c > = 0.12$, $< k > = 98.72$, D = 3, $< l > = 1.97$.