

Gustavo Henrique Pinto Soares de Barros

**Proposta de um meta-modelo para avaliação de  
robustez de redes de computadores com base  
na combinação de métricas topológicas**

Recife

2017

Gustavo Henrique Pinto Soares de Barros

**Proposta de um meta-modelo para avaliação de robustez  
de redes de computadores com base na combinação de  
métricas topológicas**

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco como requisito para obtenção do título de Bacharel em Ciência da Computação.

Universidade Federal Rural de Pernambuco - UFRPE  
Departamento de Estatística e Informática - DEINFO  
Bacharelado em Ciência da Computação

Orientador: Danilo Ricardo Barbosa de Araújo

Recife

2017



MINISTÉRIO DA EDUCAÇÃO E DO ESPORTE  
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO (UFRPE)  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

<http://www.bcc.ufrpe.br>

**FICHA DE APROVAÇÃO DO TRABALHO DE CONCLUSÃO DE CURSO**

Trabalho defendido por Gustavo Henrique Pinto Soares de Barros às 16 horas do dia 06 de setembro de 2017, na sala 02 do CEAGRI – 02, como requisito para conclusão do curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, intitulado **Proposta de um meta-modelo para avaliação de robustez de redes com base na combinação de métricas topológicas**, orientado por Danilo R. B. de Araújo e aprovado pela seguinte banca examinadora:

Danilo R. B. de Araújo  
DEINFO/UFRPE

George Gomes Cabral  
DEINFO/UFRPE

Carlos Julian Menezes Araújo  
DEINFO/UFRPE

# Lista de ilustrações

Figura 1 – Representação de redes de computadores com diferentes distribuições de <i>links</i> . Fonte: O autor. . . . .	12
Figura 2 – Representação de um neurônio artificial genérico. Fonte: O autor. . .	22
Figura 3 – Modelo genérico de uma RNA. Fonte: O autor. . . . .	24
Figura 4 – Ilustração de algumas redes reais submetidas ao método proposto. Fonte: (KNIGHT et al., 2011) . . . . .	33
Figura 5 – $\overline{EQM}$ dos limiares $f_{90}\%$ para o modelo escolhido para ambas simulações de falhas (direcionada e aleatória). Fonte: O autor. . . . .	37
Figura 6 – Desenvolvimento do EQM no decorrer das iterações. Fonte: O autor.	37

# Lista de tabelas

Tabela 1 – Métricas topológicas aplicadas na pesquisa e suas representações. . .	34
Tabela 2 – Matriz de correlação das métricas topológicas da Tabela 1. . . . .	35
Tabela 3 – Subconjuntos de métricas topológicas formados após a análise da matriz de correlação. Ver Tabela 2 . . . . .	36
Tabela 4 – Métricas de robustez e suas respectivas margens de erro 20 redes ópticas selecionadas do <i>dataset</i> (KNIGHT et al., 2011) para limiares $f_{90\%}$ direcionado. Fonte: O autor. . . . .	39
Tabela 5 – Métricas de robustez e suas respectivas margens de erro 20 redes ópticas selecionadas do <i>dataset</i> (KNIGHT et al., 2011) para limiares $f_{90\%}$ aleatório. . . . .	40

# Lista de abreviaturas e siglas

RNA	<i>Rede Neural Artificial</i>
EQM	<i>Erro Quadrático Médio</i>
$\overline{\text{EQM}}$	<i>Média do Erro Quadrático Médio</i>
MLP	<i>Multilayer Perceptron</i>
DFT	<i>Discrete Fourier Transform</i>
RAM	<i>Random Addition Model</i>
GMSW	<i>Greedy Model with Small World</i>
ms	<i>milissegundos</i>
IOT	<i>Internet of Things</i>
DASM	<i>Directed Angulation toward the Sink Node Model</i>
ASCN	<i>Assembly Supply Chain Network</i>
RWA	<i>Routing and Wavelength Assignment</i>

# Lista de símbolos

$r^{\tau}$	Indicador de Robustez
$C_G$	Condutância
$R_G$	Resistência
$f_c$	Limiar Crítico
$f_{90\%}$	Limiar 90%
$\bar{\lambda}$	Conectividade Natural
$\lambda_i$	i-ésimo Autovalor da matriz Laplaciana
$\rho$	Coefficiente produto-momento de Pearson
$\sigma$	Função Sigmóide

# Sumário

	<b>Abstract</b> . . . . .	<b>10</b>
<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>11</b>
<b>1.1</b>	<b>Visão Geral</b> . . . . .	<b>11</b>
<b>1.2</b>	<b>Problema de Pesquisa</b> . . . . .	<b>14</b>
<b>1.3</b>	<b>Objetivos</b> . . . . .	<b>14</b>
1.3.1	Objetivo Geral . . . . .	14
1.3.2	Objetivos Específicos . . . . .	14
<b>1.4</b>	<b>Etapas da Pesquisa</b> . . . . .	<b>15</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	<b>16</b>
<b>2.1</b>	<b>Análise de Robustez</b> . . . . .	<b>16</b>
<b>2.2</b>	<b>Métricas Topológicas</b> . . . . .	<b>17</b>
2.2.1	Limiar . . . . .	17
2.2.2	Assortatividade . . . . .	18
2.2.3	Condutância . . . . .	18
2.2.4	Conectividade Algébrica . . . . .	19
2.2.5	Conectividade Natural . . . . .	19
2.2.6	Entropia . . . . .	19
2.2.7	Entropia Laplaciana . . . . .	20
2.2.8	Indicador de Robustez . . . . .	20
<b>2.3</b>	<b>Correlação</b> . . . . .	<b>20</b>
<b>2.4</b>	<b>Regressão Numérica</b> . . . . .	<b>21</b>
<b>2.5</b>	<b>Redes Neurais Artificiais (RNA)</b> . . . . .	<b>22</b>
<b>2.6</b>	<b>Backpropagation</b> . . . . .	<b>24</b>
<b>3</b>	<b>TRABALHOS RELACIONADOS</b> . . . . .	<b>25</b>
<b>4</b>	<b>METODOLOGIA</b> . . . . .	<b>29</b>
<b>4.1</b>	<b>Linguagem e Ambiente de Desenvolvimento</b> . . . . .	<b>29</b>
<b>4.2</b>	<b>Visão Geral</b> . . . . .	<b>29</b>
<b>4.3</b>	<b>Seleção de Métricas</b> . . . . .	<b>29</b>
<b>4.4</b>	<b>Algoritmos Implementados</b> . . . . .	<b>30</b>
4.4.1	Simulações . . . . .	30
4.4.2	Estratégias de ataque . . . . .	30
4.4.3	Redes Neurais Artificiais . . . . .	31

4.4.4	Algoritmo de Treinamento . . . . .	31
4.4.5	Algoritmo de Teste . . . . .	32
4.5	<b><i>Dataset</i></b> . . . . .	32
5	<b>RESULTADOS</b> . . . . .	34
5.1	<b>Seleção de Métricas</b> . . . . .	34
5.2	<b>Treinamento/Validação da RNA</b> . . . . .	36
5.3	<b>Aplicação da RNA treinada</b> . . . . .	38
6	<b>CONCLUSÕES</b> . . . . .	41
6.1	<b>Artigos Publicados</b> . . . . .	41
6.2	<b>Trabalhos Futuros</b> . . . . .	41
	<b>REFERÊNCIAS</b> . . . . .	43

# Resumo

Uma crescente demanda por robustez e resiliência em redes de computadores surge da grande diversidade de aplicações das mesmas. Os sistemas modernos apresentam cada vez mais um caráter crítico, e a ocorrência de perturbações podem acarretar em sérias perdas, sejam elas humanas, econômicas ou ambientais. A fibra óptica atua nos sistemas de comunicação atuais como o principal meio de transferência de dados. Dentre suas diversas aplicações, que dependem fortemente de sua infra-estrutura, estão a Internet, televisão a cabo e sistemas que necessitam de altas taxas de transmissão. A natureza não homogênea e complexa das topologias destas redes determinam os altos custos nas avaliações das mesmas. E por este motivo, as redes ópticas são o objeto de estudo deste trabalho. A quantificação de confiança de uma rede de computadores é frequentemente alcançada através de simulações de falhas em nós e conexões. Simulações estas em que o custo temporal e financeiro crescem proporcionalmente ao tamanho da rede. Este trabalho avalia a possibilidade da obtenção de valores de métricas de robustez em redes complexas de computadores, as quais seriam adquiridas originalmente a partir de simulações, utilizando um método alternativo que aplica regressão e que tem como entrada um conjunto de valores de métricas simples correlacionadas não obtidas por simulações e utiliza um modelo de redes neurais artificiais para prever os resultados das simulações visando alcançar um menor custo temporal. Os resultados são avaliados após uma comparação entre os valores obtidos pelo modelo proposto e os valores obtidos por simulações de falhas aleatórias e simulações de ataques direcionados. Eles indicam que o método descrito apresenta uma margem de erro satisfatória, entre  $10^{-3}$  e  $10^{-9}$ , e que se atingiu com sucesso a obtenção dos valores simuláveis através do métodos de regressão em um menor intervalo de tempo.

Palavras-chave: Robustez, Resiliência, Redes Complexas, Redes Ópticas, Regressão, Redes Neurais Artificiais.

# Abstract

*A growing demand for resilience and robustness in the field of computer networks rises from the great diversity of its applications. The modern systems display an increasing critical nature, and the occurrence of perturbations may cause significant losses either human, monetary or environmental. Optical fiber acts on the current systems as the main mean of transportation. Among its variety of applications, which are heavily dependant on its infrastructure, some of them are the internet, cable television and high transmission rates systems. The non-homogeneous and complex topology nature of these networks determine their increasing avaluation cost. For these reasons, optical networks are the study object of this research. Quantifying the robustness of networks is usually accomplished by nodes and links failure simulations, on which the monetary and temporal cost scales proportionally to the network size. This research analyzes the possibility of obtaining values of robustness metrics in complex networks which would originally be obtained from simulations through an alternative regression method. This method has as inputs the values of simple metrics which are obtained through applications other than simulations and uses artificial neural networks to forecast simulation results in a smaller period. The results are obtained through a comparison between the proposed model output and the node and link failure simulation output. They indicate that the proposed model presents a satisfactory error margin, between  $10^{-3}$  and  $10^{-9}$ , thus the simulation values were reached successfully through regression on a smaller time period.*

*Keywords: Robustness, Resilience, Complex Networks, Optical Networks, Regression, Artificial Neural Networks.*

# 1 Introdução

## 1.1 Visão Geral

As redes de computadores são a base em que vários serviços são fornecidos em diversas áreas de aplicação como comércio, educação, mobilidade urbana e saúde, por exemplo (ALENAZI MOHAMMED JF; STERBENZ, 2015; WANG, 2015). Em algumas de suas utilizações, a rede é classificada como um ponto crítico para a segurança, o que significa que seu mau funcionamento pode trazer perigo à vida humana, provocar uma perda econômica substancial ou causar danos extensivos ao meio ambiente (KNIGHT, 2002). Além da função de suporte, a aplicação dos conceitos de redes são de grande importância pois eles podem ser utilizados em uma variedade de outros sistemas, tanto na natureza quanto na sociedade (WANG, 2015; JAMAKOVIC A.; UHLIG, 2007).

Cada rede possui suas próprias características quanto à tecnologia, velocidade e nichos de mercado que devem ser compatíveis com as necessidades dos serviços fornecidos pelas mesmas. A largura de banda e a latência são parâmetros críticos necessários para caracterização da qualidade dos serviços de uma rede. Vários meios físicos podem ser usados para realizar a transmissão de dados. Cada meio tem seu próprio nicho em termos de largura de banda, retardo, custo e facilidade de instalação e manutenção (RAMASWAMI RAJIV; SIVARAJAN, 2009).

A ocorrência de falhas ou ataques em alvos específicos pode causar grandes perturbações à disponibilidade dos serviços (ALENAZI MOHAMMED JF; STERBENZ, 2015), causando uma grande atração de pesquisa para a área de falhas (LUO, 2012; CHANG LIANG; WU, 2011). No cenário de falhas estudado neste trabalho ocorre um distúrbio inicial ou um evento que causa a derrubada na disponibilidade de um ou mais nós, o que causa uma mudança automática na carga designada para os vizinhos, que na maioria dos casos é capacitado para suportar a carga extra que agora é de sua responsabilidade. Entretanto, em alguns casos, estes nós vizinhos já estão sobrecarregados, e precisam redistribuir a carga para seus vizinhos, o que se repete até chegar em um ponto onde não é possível a redistribuição e eventualmente acarreta uma série de falhas nos nós, também conhecida como falha em cascata (LUO, 2012).

Sabendo da diversidade de aplicações das redes de computadores e do quão dependente a sociedade é de suas aplicações, é provável que futuramente ocorra um aumento drástico no número de sistemas que tem a infraestrutura de redes como ponto crítico de segurança, o que já vem ocorrendo gradualmente com a modernização dos sistemas (KNIGHT, 2002). Ao identificar as conexões (*links*) e nós (*nodes*) mais

vulneráveis (pontos críticos) é possível determinar a robustez da rede avaliando as suas condições sob falhas em cascata (CHANG LIANG; WU, 2011).

Existem 2 conceitos que são importantes para a quantificação da confiabilidade de uma rede: resiliência e robustez. A resiliência geralmente se refere à habilidade de uma rede de retornar a um estado prévio após sofrer algum tipo de distúrbio enquanto a robustez se refere à medida de *stress* que pode ser absorvida antes da ocorrência de uma falha (WANG, 2015). Existem alguns diferentes fatores que determinam a robustez de uma rede, quando se leva em consideração apenas fatores topológicos existem fatores como redundância ou extensão das conexões (WANG, 2015) (JAMAKOVIC A.; VAN MIEGHEM, 2008).

Pode-se observar na Figura 1 os conceitos de robustez mencionados anteriormente.

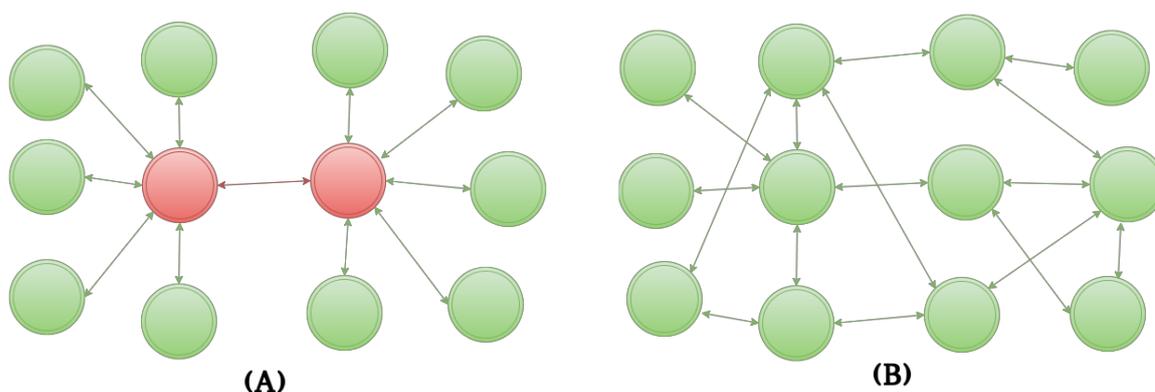


Figura 1 – Representação de redes de computadores com diferentes distribuições de *links*. Fonte: O autor.

Na Figura 1 (A) existe uma grande concentração de tráfego em um único *link*, que está destacado para facilitar sua identificação, o grupo de nós mais à direita da rede e os nós do grupo mais a esquerda da rede podem apenas se comunicar através de um único caminho, e a ocorrência de uma falha neste *link* causará a divisão da rede em 2 grandes conjuntos desconectados. Já na rede da Fig. 1 (B) existe uma diversidade maior de caminhos possíveis para realizar conexão entre os nós da rede. Além do caminho único na rede da Fig. 1 (A), observa-se também uma grande centralização nos 2 nós destacados. A ocorrência de falhas nestes nós causaria uma desintegração na comunicação de uma grande parcela da rede. Visualmente observa-se que a rede da Fig. 1 (B) suporta uma quantidade maior de *stress* quando comparada à rede da Fig. 1 (A).

Entretanto, as redes não podem ser classificadas como robustas apenas visualmente. Assim, para quantificar robustez são utilizadas métricas que são caracterizadas por certos aspectos da topologia dos sistemas estudados. E como consequência da

existência de uma variedade de fatores que caracterizam a robustez de uma rede, são conhecidas várias métricas que analisam aspectos particulares (JAMAKOVIC A.; VAN MIEGHEM, 2008). Por exemplo, a métrica utilizada por (JAMAKOVIC A.; UHLIG, 2007) (JAMAKOVIC A.; VAN MIEGHEM, 2008), que é a conectividade algébrica, pode ser utilizada para fazer a quantificação da robustez de ambas as redes da Figura 1. Como a conectividade algébrica tem como característica apresentar um valor de robustez mais elevado para as redes que possuem uma maior quantidade de caminhos independentes, pode-se confirmar a classificação realizada visualmente através desta métrica, já que o valor da Conectividade Algébrica para a rede da Fig. 1 (A) é de 0,25 e para a rede da Fig. 1 (B) é de 0,41. Mas nem sempre é fácil realizar um estudo de robustez, pois ao selecionar as métricas podem ocorrer divergências de resultados, em que para um determinado cenário uma métrica afirma que uma rede estudada possui maior robustez em relação a outra, e para o mesmo cenário, outra métrica diz o contrário. Um exemplo desta ocorrência pode ser estudada em Wang (2015) em que se aplica um estudo de robustez utilizando as métricas condutância e indicador robustez e ao analisar os resultados o autor se depara com uma contradição. Visando eliminar a possibilidade de tais divergências e facilitar os estudos de resiliência novas abordagens precisam ser consideradas.

Em De Araújo et al. (2015) observa-se uma abordagem de aplicação de técnicas de regressão numérica para aprendizado de máquinas, aplicada para previsão de desempenho em redes de computadores. Nela se utiliza um grande conjunto de treinamento para geração de um modelo adaptativo para posterior previsão de uma ou mais variáveis a partir de um vetor de entrada.

Um dos mais simples modelos de regressão é o que envolve apenas uma combinação linear entre um número fixo de funções de *input*. Entretanto por ser muito simples existem diversas limitações para o mesmo, sendo necessárias adaptações das funções base, que são funções em que a partir delas pode-se abranger todo o conjunto de funções ao qual ela pertence, para ser aplicado em problemas de larga escala. Uma abordagem alternativa mais avançada é manter fixo o número de funções base, entretanto permitir que tais funções sejam adaptativas, ou seja, permitir que os valores de parâmetro das funções se adaptem durante o processo de treinamento. O modelo que aplica este método mais conhecido é o perceptron multicamadas (*multilayer perceptron*) que é classificado como um tipo de rede neural (BISHOP, 2007).

Uma análise completa de robustez da rede exige o uso de diversas métricas combinadas como, por exemplo em Wang et al (2015) em que ele aplica um conjunto de 4 métricas topológicas a condutância, indicador de robustez e limiares, incluindo a simulação de falhas em nós e conexões das redes de computadores para avaliação de algumas destas métricas, o que é uma tarefa computacionalmente custosa. Portanto, este

trabalho foca na proposição de um indicador simples, que possa fornecer uma medida global de robustez da rede de computador e que possua baixo custo computacional para o seu cálculo, o que possui grande relevância para a área.

As redes ópticas hoje são as principais responsáveis pelo atendimento da crescente demanda por taxa de transmissão, atuando como redes de transporte responsáveis por trafegar os dados agregados de vários usuários (RAMASWAMI RAJIV; SIVARAJAN, 2009). Devido à sua importância e à gravidade da ocorrência de falhas em redes de transporte ópticas, este tipo de rede será usado como estudo de caso para as análises de robustez deste trabalho.

## 1.2 Problema de Pesquisa

As pesquisas que tem como objetivo a determinação da robustez escolhem métricas que, de acordo com os autores, se completam ou de alguma forma preenchem a relação de características necessárias. Estas características como, por exemplo, quantidade de caminhos redundantes devem ser vistas pelo autor do estudo como suficientes para expressar a robustez do ambiente em questão a ser analisado (ALENAZI MOHAMMED JF; STERBENZ, 2015; WANG, 2015; JAMAKOVIC A.; UHLIG, 2007).

Neste sentido, o problema de pesquisa abordado neste trabalho pode ser enunciado por meio da seguinte pergunta: "É possível obter um indicador global de robustez de redes combinando diferentes medidas de robustez, escolhidas por uma técnica de correlação dentro de um grande conjunto de possíveis métricas presentes na literatura, por meio de métodos regressivos baseados em aprendizagem de máquina?"

## 1.3 Objetivos

### 1.3.1 Objetivo Geral

Propor e analisar viabilidade um preditor de robustez de redes baseado em métodos de regressão numérica que combinam métricas topológicas que possuem baixo custo computacional disponíveis na literatura selecionadas por métodos de correlação.

### 1.3.2 Objetivos Específicos

- Analisar através de um estudo comparativo a correlação entre métricas de robustez selecionadas da literatura quando aplicadas a redes ópticas implantadas em diferentes países.

- Analisar o panorama geral da robustez de redes reais e apresentar os resultados obtidos da aplicação do regressor proposto pelo trabalho.

## 1.4 Etapas da Pesquisa

Este trabalho foi conduzido em diversas etapas, que podem ser resumidas pelos itens a seguir:

- Foi realizada continuamente uma busca por novas referências na literatura durante as etapas iniciais da pesquisa a fim de manter uma percepção atualizada sobre o cenário da área em que o trabalho se insere. Assim como buscar novas métricas para quantificação de robustez topológica e novos métodos para avaliação de redes de computadores visando obter novas perspectivas e entendimento mais profundo sobre o problema tratado.
- Definiu-se experimentos e técnicas a serem utilizadas tendo como base os conceitos já pré-definidos na etapa de projeto, são eles: utilização da linguagem de programação Java no ambiente eclipse, utilização de técnicas de regressão numérica em redes neurais artificiais mais especificamente utilização do perceptron multicamada e utilização do coeficiente de correlação de Pearson para escolha métricas.
- Criou-se um *dataset* adequado às necessidades do projeto para treinamento das redes neurais, o *dataset* criado terá como base um conjunto de redes reais disponíveis em (KNIGHT et al., 2011).
- Realizou-se a execução dos experimentos com base nas definições das etapas anteriores buscando obter a partir desta etapa os valores numéricos das métricas para em seguida correlaciona-los para posteriormente aplicar a regressão.
- Foram realizadas análises dos resultados visando a aplicação de métodos avaliadores tradicionais, e se possível, buscar novos métodos de avaliação e aplicação destes métodos a fim obter um resultado com credibilidade.
- Elaborou-se a monografia descrevendo com a maior riqueza de detalhes possíveis os métodos utilizados, resultados obtidos e avaliações realizadas. Chegando assim ao fim das etapas desta pesquisa.

## 2 Fundamentação Teórica

Este capítulo fornece uma base mínima de conhecimento necessário para o entendimento deste trabalho. Iniciando com conceitos que definem a ação de realizar um estudo de robustez, passando pelos métodos de correlação aplicados e finalizando com os conceitos de regressão e redes neurais artificiais.

Já foi mencionado anteriormente neste trabalho a importância e variedade das redes de computadores, mas como se faz a medida da diferença entre estas redes?

### 2.1 Análise de Robustez

A análise de robustez fornece uma abordagem para estruturar e quantificar as situações em que a incerteza é alta como, por exemplo, situações críticas em que é necessário um determinado grau de controle sobre os recursos disponíveis (KNIGHT, 2002). Além da função de suporte, a aplicação dos conceitos de redes são de grande importância pois eles pode. A forma mais simples de determinação de robustez é fazer uma análise dos seus valores numéricos de suas propriedades topológicas como, por exemplo, analisar a presença de redundância dentre os possíveis caminhos de uma rede. Assim, se for considerado dois nós quaisquer em uma rede, um deles sendo a origem e ou outro o destino, podem existir uma variedade de caminhos possíveis para se chegar ao destino. Quando uma ou mais rotas falham, é ideal que os nós origem e destino ainda consigam se comunicar, e em alguns casos é essencial que essa comunicação por uma rota alternativa não sofra nenhuma perda em relação ao desempenho ou disponibilidade.

Assim, quando se deseja construir uma rede projetada para ser robusta, geralmente se aplicam modelos que possuem uma grande densidade, reforçando a mesma com diversos caminhos redundantes. Alternativamente, pode-se projetar redes com foco na resiliência, que são altamente adaptativas e possuem habilidade de se recuperar de perdas da capacidade de transporte. As redes resilientes são mais economicamente viáveis quando comparadas com as robustas, entretanto são mais difíceis de serem implementadas, já que as redes resilientes são definidas pelos religamentos dos *links* após alguma perturbação na mesma e esses religamentos por sua vez são mais complexos. Já as redes robustas, não são adaptáveis e tem como foco simplesmente alta densidade e redundância (TRAN; DOMERÇANT; MAVRIS, 2016).

Levando isto em consideração, observa-se que nas redes de computadores, a robustez, escalabilidade e *performance* de protocolos e aplicações, todos dependem da

topologia aplicada nas mesmas. Consequentemente, a análise da topologia de uma rede, foca primariamente nas propriedades mais complexas das mesmas como, por exemplo, propriedades espectrais do grafo que representa a rede. Com o desenvolvimento desta prática ao longo dos anos, resultou na definição de uma grande variedade de importantes métricas qualificadas para caracterizar quantitativamente determinados aspectos dos sistemas estudados (JAMAKOVIC A.; VAN MIEGHEM, 2008).

## 2.2 Métricas Topológicas

As métricas topológicas diferem umas das outras em diversos aspectos, como relevância atribuída aos aspectos topológicos estudados, método de cálculo e custo computacional por exemplo. Algumas métricas são mais simples de serem calculadas quando comparadas a outras. O conceito de métricas simples aplicado nesta pesquisa se refere a métricas que utilizam poucos recursos computacionais para serem calculados, por exemplo, quando comparamos a conectividade algébrica de Jamakovic et al (2007, 2008) que é obtida pelo cálculo dos autovalores da matriz que representa um grafo, e que possui duração da grandeza de milissegundos para todas as redes do *dataset* utilizado neste trabalho, com a métrica do limiar 90%, obtida por simulação, e que tem duração na grandeza de segundos para o mesmo conjunto de topologias estudadas, observa-se que existe uma grande diferença de recursos utilizados. Além disso, o custo da aplicação de simulação cresce surpreendentemente com o crescimento do número de nós e *links* assim como a complexidade dos cenários de aplicativos (RAMASWAMI RAJIV; SIVARAJAN, 2009).

As métricas aplicadas nesta pesquisa podem ser divididas em 2 grupos, as métricas obtidas a partir de simulações (referidas neste trabalho como métricas simuladas) e as métricas obtidas a partir de cálculos utilizando os meta-dados das redes (referidas nesta pesquisa como métricas não-simuladas). A métrica simulada é o Limiar 90% (*Threshold 90%*). As métricas não simuladas são: Conectividade Algébrica (*Algebraic Connectivity*), Assortatividade (*Assortativity*), Condutância (*Conductance*), Entropia (*Entropy*), Entropia Laplaciana (*Laplacian Entropy*), Conectividade Natural (*Natural Connectivity*) e Indicador de Robustez (*Robustness Indicator*).

### 2.2.1 Limiar

O limiar é uma métrica obtidas a partir de simulações de falhas, o seu conceito está relacionado à quantificação da fração de nós que devem ser removidos em uma rede até que seu maior componente conectado possua um tamanho igual a fração da rede original predeterminada.

A condição de parada para obtenção desta métrica é quantificada pela aplicação

do conceito de maior componente conectado. O componente conectado é um termo da teoria de grafos que expressa a ideia que dados 2 nós quaisquer, eles farão parte do mesmo componente se existe um ou mais caminhos entre eles (BISHOP, 2007).

O limiar  $f_{90\%}$  representa a fração de nós que devem ser removidos para que o maior componente conectado da rede contenha 90% do tamanho original. Nesta variação do limiar são aplicadas duas estratégias de simulação de falhas, a aleatória e a direcionada (LEWIS, 2009; WANG, 2015).

### 2.2.2 Assortatividade

A assortatividade pode ser descrita como a preferência de um determinado nó de uma rede por se conectar a outros nós semelhantes em alguma característica. E apesar de que essa característica de similaridade possa variar, muitos pesquisadores geralmente associam a assortatividade a termos de grau de nós. Além disso, a Assortatividade também é comumente correlacionada entre 2 nós e existem formas diferentes de capturar tal correlação. As mais conhecidas são o coeficiente de assortatividade e a conectividade de vizinhos (LEUNG; CHAU, 2007).

O coeficiente de assortatividade é a aplicação do coeficiente de correlação de Pearson do grau de um par de nós conectados. Valores positivos do coeficiente assortatividade entre 2 nós representam uma correlação entre nós de mesmo grau, enquanto valores negativos representam correlações entre nós de graus diferentes (LEUNG; CHAU, 2007).

### 2.2.3 Condutância

A Condutância, vista em (WANG, 2015), é uma versão normalizada da métrica Resistência, que captura a robustez de uma rede incorporando o número de caminhos paralelos a partir de uma determinada origem para o mesmo destino, ou seja, redundância e também o comprimento entre cada par de nós.

A Resistência e  $R_{ij}$  entre um par de nós  $i$  e  $j$  pode ser definida como a diferença de potencial entre os mesmos quando a unidade medida é injetada no nó  $i$  e retirada no nó  $j$ . A Resistência efetiva  $R_G$  de um grafo, é a soma de todos os  $R_{ij}$  entre todos os pares de nós da rede. Um eficiente método do cálculo da Resistência efetiva de um grafo em termos de autovalores é:

$$R_G = N \sum_{i=1}^{N-1} \frac{1}{\mu_i}. \quad (1.1)$$

Este trabalho utiliza uma versão normalizada do valor da Resistência, também conhecida como a Condutância efetiva do grafo e pode ser calculada através da seguinte

formula:

$$C_G = \frac{N - 1}{R_G}. \quad (1.2)$$

### 2.2.4 Conectividade Algébrica

A Conectividade Algébrica de (JAMAKOVIC A.; VAN MIEGHEM, 2008) é considerada uma propriedade espectral do grafo e tem um significado relevante quando se estuda a topologia de um grafo que, no contexto deste trabalho, representa a rede. Ela tem como característica apresentar um valor de robustez mais elevado para as redes que possuem uma maior quantidade de caminhos independentes. Analisando a conectividade algébrica, pode-se determinar algumas propriedades que um grafo possui, tal como saber se o mesmo está desconectado apenas observando se o seu valor é zero. Paralelamente, observa-se que é possível obter mais informações sobre o grafo como a multiplicidade de zeros dentre os autovalores do mesmo por este valor ser igual a quantidade de componentes desconectados do mesmo. A partir destas observações é admissível chegar a algumas conclusões sobre a conectividade algébrica, a primeira é que quanto maior for o seu valor, maior será o grau de dificuldade para separar o grafo em componentes independentes e ainda é possível quantificar a extensão de que um grafo pode acomodar falhas de nós e conexões.

### 2.2.5 Conectividade Natural

Jun et al. (2010) propõe esta métrica que tem como característica o aumento monótono de seu valor quando novos *links* são adicionados, ou seja, um grafo com 0 conexões tem o seu menor valor possível da conectividade natural, enquanto um gráfico com o máximo número de conexões tem o seu valor máximo. A métrica pode ser calculada através da seguinte fórmula:

$$\bar{\lambda} = \ln \left( \frac{1}{N} \sum_{i=1}^N e^{\lambda_i} \right), \quad (2)$$

nos quais  $\lambda_i$  representa o  $i$  – ésimo autovalor da matriz Laplaciana e  $N$  é número de nós.

### 2.2.6 Entropia

O termo entropia tem diversas interpretações que dependem do contexto em que se insere. No contexto deste trabalho, esta métrica pode ser descrita como a entropia do grau de nós, que é uma propriedade estatística fundamental que fornece uma

caracterização do comportamento de redes dinâmicas e sua capacidade de permanecer funcional quando submetida a perturbações aleatórias. Existem relatos que afirmam a existência de uma correlação positiva entre a entropia e estabilidade de uma rede, medida através oscilação da taxa de declínio após perturbações aleatórias (DEMETRIUS LLOYD; MANKE, 2005).

### 2.2.7 Entropia Laplaciana

O conceito de Entropia Laplaciana aplicado nesta pesquisa se refere a entropia da Transformação Discreta de Fourier (DFT) dos autovalores Laplacianos de um determinado grafo que está descrito em (ARAÚJO DANILO RB; BASTOS-FILHO, 2014). Esta métrica fornece informações precisas sobre o tipo de estrutura apresentada pela topologia da rede. Baseado em seu valor, representado por  $I(\hat{\mathcal{F}})$ , pode-se identificar a presença de características dos modelos *small-world* ou *scale free networks* e ainda a ausência de *hubs*.

### 2.2.8 Indicador de Robustez

Esta métrica, descrita em (WANG, 2015), tem como característica a medida de robustez em termos de quantidade de caminhos redundantes em uma topologia de rede dividido pelo número total de nós da mesma. Ela pode ser calculada utilizando a fórmula:

$$r^{\tau} = \frac{\mu - L^m}{N_S}, \quad (3)$$

em que  $r^{\tau}$  representa o valor do Indicador de Robustez,  $\mu$  representa o número ciclomático que calcula o total de caminhos alternativos,  $N_S$  representa o número total de nós e  $L^m$  representa o número de múltiplos *links* entre 2 nós.

## 2.3 Correlação

A correlação é uma medida estatística que indica a extensão em que duas ou mais variáveis avançam juntas. Uma correlação positiva indica que estas variáveis crescem ou decrescem paralelamente, já uma correlação negativa indica que enquanto uma variável cresce a outra decresce. Um coeficiente de correlação é uma medida estatística que quantifica este fenômeno e permite que seja possível prever mudanças no valor das outras variáveis dado que se tenha a mudança em um dos valores envolvidos.

No entanto, quando se consegue prever confiavelmente as mudanças em uma variável a partir da outra, geralmente surge uma tendência de se pensar que a mudança

de uma variável causa a mudança da outra, o que na realidade não acontece. Ou seja, a correlação não implica causalidade.

Existem várias métricas de quantificação da correlação entre duas variáveis. Dentre elas, a mais comum é o coeficiente de correlação produto-momento de Pearson, ele dimensiona linearmente a correlação entre duas variáveis fornecendo um resultado entre +1 e -1 em que +1 corresponde a correlação linear positiva total, 0 corresponde a nenhuma correlação e -1 correlação linear negativa total.

Digamos então que se deseja quantificar a correlação entre dois conjuntos  $\{x_1, \dots, x_n\}$  que contém  $n$  valores e  $\{y_1, \dots, y_n\}$  que também possui  $n$  valores, então, o coeficiente de Pearson ( $\rho$ ) é calculado utilizando a seguinte fórmula:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (4.1)$$

onde:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4.2)$$

e

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (4.3)$$

## 2.4 Regressão Numérica

A regressão numérica é um método estatístico em que se obtém uma estimativa da relação entre variáveis através da combinação de diversas técnicas. Mais especificamente, a aplicação da regressão proporciona auxílio para o entendimento de como um valor de uma variável dependente se comporta quando uma ou mais das variáveis independentes se altera. Em um cenário típico de regressão, pode-se obter como saída do processo uma medida quantitativa tal como o limiar 90% como de Araújo et al (2015) descreve. Em geral esse tipo de processo se baseia em conjunto de características, tais como as métricas topológicas e medidas de elementos físicos da rede de computadores estudada. Neste trabalho, é feito o treinamento de uma rede neural artificial para executar regressão e obter valores do limiar 90%.

Uma importante tarefa para otimizar a qualidade do modelo de regressão aplicado é a de reduzir o conjunto de variáveis de decisão a um subconjunto que consiga representar o problema abordado completamente visando evitar o *overfitting*. O *overfitting* é um problema que ocorre em modelos de regressão, em que se induz a obtenção de conclusões enganosas. Para evitar o *overfitting*, nesta pesquisa utiliza-se a validação cruzada. Este passo adicional permite a obtenção de melhores propriedades generalizadoras e facilita a interpretação da função obtida (ARAUJO DANILO RB; BASTOS FILHO, 2015a).

## 2.5 Redes Neurais Artificiais (RNA)

As redes neurais artificiais são modelos matemáticos que são compostos por um conjunto de unidades computacionais básicas, que no contexto de representação de redes através de grafos seria equivalente a um nó. A nomenclatura rede neural é inspirada pelas operações biológicas realizadas por células especializadas chamadas de neurônios. Um neurônio artificial é uma célula que possui diversas entradas e pode ser ativado por um processo externo. Uma rede é formada a partir do encadeamento de conexões entre estes nós ou neurônios (HAYKIN, 2010).

Dependendo da quantidade de ativações imposta pelo agente de controle externo, o neurônio artificial pode passar a produzir sua própria ativação e enviar os resultados através de suas saídas. Assim, um *input* específico ou caminhos específicos são fortalecidos de acordo com seu desempenho e passam a possuir pesos maiores em relação aos seus demais semelhantes. A hipótese é que já que o cérebro humano é uma rede de neurônios, é possível imita-lo através de uma modelagem dos neurônios e conectando-os via grafo de pesos. Na Figura 2 pode-se observar um modelo genérico de um neurônio artificial (HAYKIN, 2010).

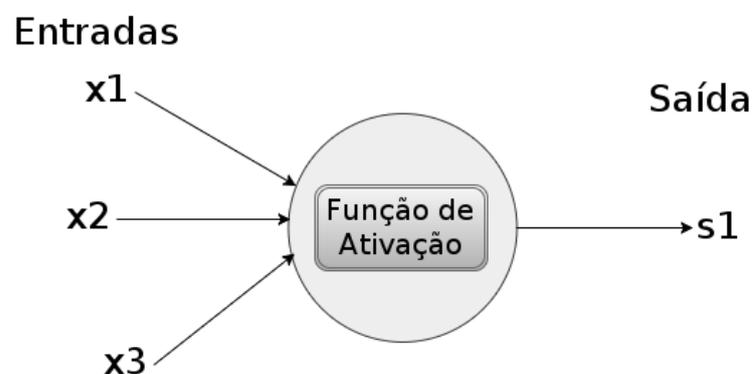


Figura 2 – Representação de um neurônio artificial genérico. Fonte: O autor.

Um neurônio recebe um conjunto de entradas com diferentes pesos, em seguida realiza o processamento destes valores usando a sua função de ativação, e transmite o resultado adiante com entrada para uma próxima camada, ou como a saída da rede. Neste trabalho, aplica-se um modelo que apresenta funções derivadas rotuladas  $Z_m$ , que são criadas a partir da combinação das entradas ( $X_p$ ) e seus respectivos pesos. As características derivadas são avaliadas para  $P$  variáveis de acordo com a seguinte fórmula:

$$Z_m = \sigma\left(\alpha_0 + \sum_{p=1}^P \alpha_p X_p\right), \quad (5)$$

em que  $m \in \{1, 2, \dots, M\}$  e  $M$  são os números de neurônios na camada escondida.  $\alpha_0$  é o *bias* do neurônio. Cada neurônio executa a combinação linear dos valores de entrada de acordo com a função de ativação. As funções de ativação não-lineares mais comuns são a sigmóide e a função tangente hiperbólica (*Hyperbolic tangent function* ou *tanh*). Neste trabalho, aplica-se a função de ativação sigmóide através da seguinte equação:

$$\sigma(v) = 1/(1 + e^{-v}), \quad (6)$$

em que  $\sigma(v)$  corresponde à aplicação da função sigmóide em uma entrada  $v$ . Também, se aplica o modelo de RNA mais usado na literatura o chamado *Multilayer Perceptron* (MLP). A Figura 3 demonstra um modelo MLP genérico utilizado para regressão. As redes neurais multi-camadas funcionam como um encadeamento de neurônios que se organizam em forma de camada. A primeira camada é responsável pela entrada da rede neural que contém um ou mais neurônios e que ao serem ativadas realizam uma determinada computação sobre os dados de entrada e passam os resultados para próxima camada. Em seguida se encontra uma ou mais camadas escondidas que são responsáveis por realizar seus próprios cálculos e encaminhar seus resultado para camada seguinte até chegar a sua última camada que contém as saída da rede.

O objetivo é obter uma rede *MLP* treinada para que posteriormente seja possível realizar a inserção de um conjunto de dados, que preferivelmente ainda não tenha sido aprendido pela rede, e seja produzido uma predição do resultado apropriado. Isso é possível porque na etapa de treinamento e teste se possui um par de valores de entrada e saída de um determinado alvo. O treinamento neste caso envolve o aprendizado correto do peso das conexões que produz a saída deseja dado uma determinada entrada. Quando finalizado o treinamento e teste, a rede e seus respectivos pesos treinados formam uma função que opera sobre os dados de entrada.

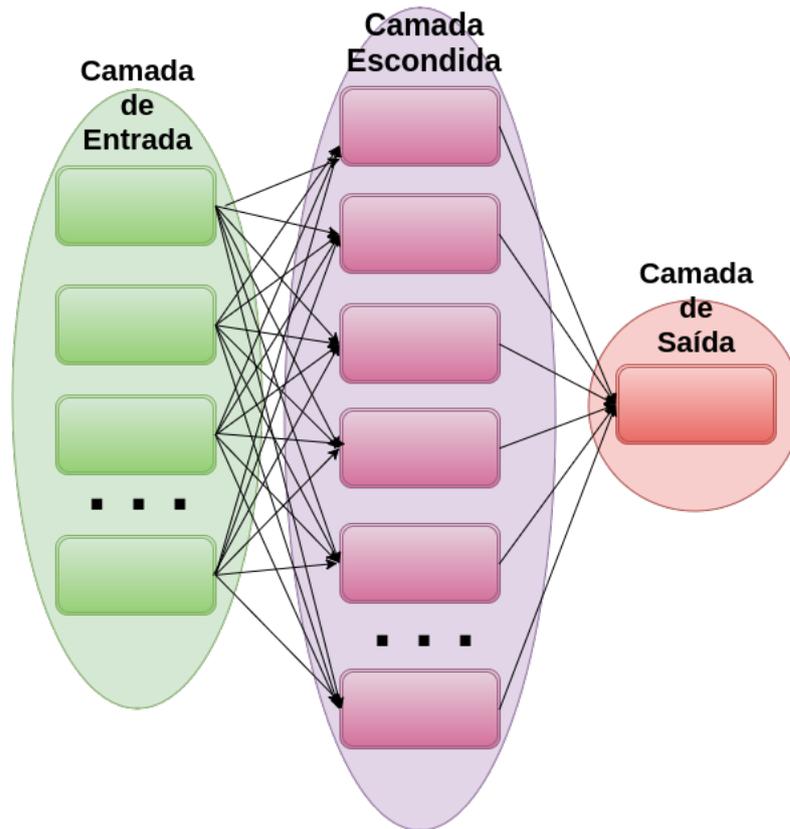


Figura 3 – Modelo genérico de uma RNA. Fonte: O autor.

## 2.6 Backpropagation

*Backpropagation* é frequentemente vista como um método de adaptação de uma RNA visando otimização das margens de erro obtidas. A etapa chamada de *backpropagation* aplicada nesta pesquisa consiste em calcular recursivamente as derivadas das funções de erro respeitando os pesos e os *bias* do modelo, através da propagação do erro, partindo da camada de saída até chegar na camada de entrada, uma camada por vez. Já que cada camada da rede neural consiste em uma simples função aplicada sobre o valor da camada anterior, é possível definir regras que determinam analiticamente as derivadas da função de erro (PAI, 2017).

### 3 Trabalhos Relacionados

Em (ALNAZI MOHAMMED JF; STERBENZ, 2015) os autores escolhem um conjunto de ambientes reais com objetivo de avaliar e melhorar a resiliência topológica. Para realizar a avaliação é escolhido e apresentado um conjunto de métricas, como o *link-betweenness* por exemplo, e aplicando-o os autores produzem um *framework* que utiliza uma noção de espaço em 2 dimensões e faz simulações de falhas para quantificar a resiliência. Utilizando tal *framework*, os autores avaliam a resiliência dos ambientes escolhidos e de posse dos resultados eles submetem as redes estudadas a um algoritmo guloso que faz a inserção de novos *links* com objetivo de aumentar a resiliência da rede e em seguida é recalculada a resiliência e os valores obtidos são comparados. Os autores chegam à conclusão de que não existe uma métrica ideal para ser a função que mede a adição de novos *links*. Entretanto, ao adicionar novos *links*, uma das métricas escolhidas, que é a “*link-betweenness*”, apresenta a maior consistência entre as métricas estudadas.

Em (JAMAKOVIC A.; UHLIG, 2007) os autores reconhecem a métrica conectividade algébrica como uma propriedade espectral do grafo que representa a rede e um importante parâmetro que está ligado a diversas análises de problemas relacionados à robustez das redes. Partindo desta premissa o autor tem como objetivo avaliar a relação entre a métrica e robustez do ambiente estudado. As conclusões obtidas foram que a conectividade algébrica é considerada como uma medida legítima de robustez em modelos de redes complexas e quanto maior for o seu valor, maior será o número de caminhos dispersos.

Em (WANG, 2015) a principal ideia dos autores é fazer um estudo que produza um conjunto de resultados que quantifiquem a robustez de diversos sistemas de metrô espalhados pelo mundo. São selecionados 33 representações de trilhos reais na forma de grafos, e aplicados métodos de análise de robustez com métricas selecionadas pelos autores através de simulações a falhas aleatórias e ataques a alvos. Os autores concluem que, ao observar algumas das métricas utilizadas, elas divergem em resultados e isso se dá por estas métricas atribuírem importância a aspectos diferentes de robustez. E que essa divergência causa uma contradição nos resultados obtidos. E os autores apontam que esta contradição mostra que é difícil fazer a determinação da robustez, e que os resultados obtidos pedem por mais focos de pesquisa na área para que o planejamento de sistemas de mobilidade urbana possa ser beneficiado futuramente.

Em (ARAUJO DANILO RB; BASTOS FILHO, 2015b) são utilizadas técnicas de aprendizagem de máquina para quantificar desempenho de redes através do conceito de probabilidade de bloqueio, que é definido como a relação entre o número de requi-

sições bloqueadas em uma rede de computadores e o número total de requisições na mesma. A metodologia aplicada considera as características da rede como, por exemplo, sua determinada topologia física, o número máximo de comprimentos de ondas disponíveis, o fator de isolamento dos *switches* ópticos, a demanda de tráfego e o algoritmo de roteamento *Routing and Wavelength Assignment* (RWA). E investiga se é possível a determinação da probabilidade de bloqueio.

Partindo desta premissa, considerando os fatores da rede mencionados, é aplicado o modelo de regressão de Redes Neurais Artificiais para obter rápida e precisamente uma estimativa para o valor da probabilidade de bloqueio. Os autores concluem que o método proposto obtém resultados com precisão similar ao método de simulação discreta de eventos e com tempo de avaliação similar ao método de expressões analíticas, e que a proposta é generalista, podendo ter maior precisão se adaptações forem realizadas.

Os autores Araújo et al. (2017) propõem um método sistemático para analisar diferentes *backbones* ópticos utilizando algoritmos de clusterização não supervisionados. Em (ARAUJO DANILO RB; MARTINS FILHO, 2017) ainda é analisada a aplicação de uma métrica que se baseia na entropia dos coeficientes de Fourier dos autovalores da Laplaciana. A pesquisa conclui que o modelo proposto que utiliza o algoritmo de aprendizagem de máquina *K-means* apresenta uma alternativa fácil para a análise das propriedades topológicas de uma rede e que a métrica estudada é melhor em termos de capacidade individual quando comparada às outras métricas.

No trabalho apresentado em (ARAUJO DANILO RB; BASTOS FILHO, 2015a) é realizada uma aplicação de metodologia similar à proposta em (ARAUJO DANILO RB; BASTOS FILHO, 2015b), usando redes neurais para prever o desempenho de redes em seis redes reais implantadas para que seja realizada a avaliação do método. Também são propostas e aplicadas quatro novas métricas relacionadas à camada física e os resultados são comparados com resultados de simulações em redes de eventos discretos. Os autores concluem que a aplicação de RNA para a estimativa dos valores da probabilidade de bloqueio é uma abordagem promissora, pois a margem de erro da estimativa é pequena. Vale salientar que nenhum estudo com foco em robustez foi realizado neste trabalho.

Em (SARKAR TUHIN; ROOZBEHANI, 2016), os autores afirmam apresentar um conjunto de novas noções de características assintóticas que acrescentam um ponto de vista diferenciado sobre o estado atual do conceito de robustez em grandes redes interconectadas a medida que as dimensões da determinada rede cresce. Mais especificamente, os autores focam em solucionar a questão de como a robustez das redes se comporta enquanto as dimensões das mesmas varia. Este trabalho tem como objeto de pesquisa um conjunto de topologias construídas de aplicações nas áreas de transporte e economia, o que se situa exatamente no conjunto de grandes redes com dimensões

dinâmicas, e que são de grande importância para a sociedade, fazendo análises das características apresentadas por estas redes que possuem escalabilidade ruim.

Além disso, em (SARKAR TUHIN; ROOZBEHANI, 2016) ainda são analisados como os impactos causados por perturbações de *links* afetam as noções de robustez em uma rede, permitindo identificar os pontos críticos nas redes que podem conduzir a um projeto de um design mais robusto.

Os autores em (SARKAR TUHIN; ROOZBEHANI, 2016) concluem que os resultados obtidos em sua pesquisa fornecem uma compreensão válida da relação entre os *links* de uma topologia e a fragilidade da respectiva rede. E que no decorrer de sua pesquisa foi possível identificar os pontos críticos em uma rede frágil e que ao modificá-los com reforço de novos *links* foi possível reconhecer uma melhora na resiliência da rede. Ainda, os autores ressaltam que através do conjunto de redes escolhidos, redes econômicas e de transporte, mostra que as ideias e conceitos de resiliência aplicados podem se estender e assim proporcionar compreensão em noções de estabilidade em outros contextos. E finalizam seu estudo apontando que existe grande potencial no estudo de matrizes com mais de um autovalor dentro do círculo de unidade.

Em (TANG, 2016) os autores tem como objeto de estudo uma rede referente a uma linha de montagem (*Assembly Supply Chain Network – ASCN*) que é composta por fabricantes geograficamente localizados em diferentes regiões. Para ser analisada a robustez da ASCN em questão quando ela sofre alguma interrupção por eventos catastróficos, é construído um modelo de falhas em cascata de propagação de riscos. E neste modelo, são considerados diferentes cenários e a partir deste modelo são desenvolvidas equações de probabilidade. Os autores usam a perda de capacidade de produção (*Production Capability Loss*) como índice de robustez. E a partir daí são conduzidas simulações de falhas para estimar a robustez, e através desta simulação, são comparados os valores de robustez obtidos em diferentes graus de intensidade das conexões e limiares de nós com intuito de encontrar pontos críticos nas conexões e o limiar máximo de nós que afeta a robustez da ASCN.

Tang (2016) chega à conclusão, através de seus experimentos, que a rede de maior complexidade, por apresentar uma maior quantidade de redundância entre outras características, apresenta uma maior robustez. Ele concluem ainda que o método utilizado por ele se provou ser uma ferramenta valiosa na análise de robustez em ASCN quando se deseja considerar ocorrência de ataques e/ou falhas. E ainda que seu método utilizado serve como base para modelos de otimização e modelos de controles de falha em cascata.

A pesquisa da referência (QIU, 2016) se integra com diversos campos da área de redes de computadores e que vem recebendo muita atenção atualmente, a Internet das coisas (*Internet of Things – IOT*). Na IOT a robustez é uma questão desafiadora, já que

ela contém múltiplos tipos de redes heterogêneas. Neste contexto, aprimorar a robustez de uma estrutura topológica, ou seja, resistir a uma certa quantidade falhas em seus componentes, é de grande importância especialmente para suas rede leves de energia limitada. Entretanto, na IOT topologias de alta performance também são necessárias. Ainda em (QIU, 2016), os autores propõem um modelo guloso com propriedades do mundo pequeno (*Greedy Model with Small World – GSMW*) destinado à redes sensores heterogêneas. Inicialmente os autores apresentam 2 critérios gulosos aplicados ao GSMW para distinguir a importância de diferentes nós dentro da rede. Em seguida é apresentado o algoritmo que transforma a rede, que já contém características do mundo pequeno (quantidade pequena de *links* de alta distância), adicionando redundâncias entre determinados nós de acordo com suas respectivas importâncias locais. E posteriormente é realizada a avaliação da performance. Os autores de (QIU, 2016) concluem que ao comparar os resultados obtidos da aplicação do modelo *small-world* com outros modelos, dentro da mesma probabilidade, o número médio do tamanho dos caminhos do *small-world* é reduzido mais rapidamente enquanto seu coeficiente de clusterização decresce mais vagarosamente, o que sugere que o GSMW proposto possui melhores propriedades do que os outros modelos. Ao avaliar a robustez, os resultados indicam que tanto nos casos de falhas gerais quanto nos casos de falhas específicas, a latência do GSMW é menor quando comparado ao modelo *Directed Angulation toward the Sink Node Model - DASM*. Para se aprofundar mais sobre este assunto, leia a referência (QIU, 2016).

Este trabalho tem o intuito encontrar um modelo baseado na aplicação de RNA que substitua o modelo original baseado em simulações que demandam grandes investimentos de tempo e recursos computacionais, aplicando correlação entre as características das métricas topológicas para avaliar se os resultados obtidos fornecem um indicador de robustez confiável, que satisfaça as necessidades de um estudo de resiliência e que apresente um menor custo computacional e temporal.

## 4 Metodologia

Este capítulo visa descrever os métodos e ferramentas utilizadas na produção desta pesquisa.

### 4.1 Linguagem e Ambiente de Desenvolvimento

Por questões de compatibilidade com materiais acessíveis e quantidade de suporte fornecido pela comunidade, a Linguagem utilizada foi o Java e o ambiente o Eclipse.

### 4.2 Visão Geral

Nesta pesquisa, deseja-se demonstrar que é possível, utilizando uma combinação de métricas topológicas simples, evitar a aplicação de simulações custosas para a determinação de robustez em redes de computadores. Para julgar esta hipótese, é preciso que diversas métricas topológicas sejam analisadas como candidatas para atuarem como entrada do modelo proposto.

A implementação do modelo proposto pode ser dividida nas seguintes etapas: seleção de métricas, treinamento/validação da RNA e aplicação da RNA treinada. As etapas possuem uma ordem precedência que é definida pela dependência que suas entradas apresentam em relação ao resultado de outras etapas. A aplicação da RNA depende do treinamento/validação prévio da mesma, que por sua vez depende da seleção de métricas que irão compor a entrada da RNA. Assim, seguindo a ordem de precedência, a primeira etapa descrita é a seleção de métricas.

### 4.3 Seleção de Métricas

O subconjunto de métricas topológicas simples utilizadas como entrada para a RNA foram selecionadas do conjunto total  $M$  de métricas que contém 7 métricas de baixo custo, são elas: conectividade algébrica, assortatividade, condutância, entropia do grau de nós,  $DFT$  da entropia laplaciana, conectividade natural e indicador de robustez. Primeiramente foi quantificada a correlação entre todas as métricas pela aplicação do coeficiente produto-momento de Pearson, gerando uma matriz de correlação  $M \times M$ . Posteriormente, cada uma das  $M$  métricas foram atribuídas como núcleo do subconjunto de entrada e selecionadas as 3 métricas com os maiores valores de correlação em relação

ao núcleo. Finalmente, para cada subconjunto formado, foram realizados treinamentos e validações das redes neurais e analisados os seus valores de erros.

## 4.4 Algoritmos Implementados

### 4.4.1 Simulações

O valores dos limiares são obtidos através de simulações. Em de Araújo et al (2015) são utilizadas 1000 chamadas de bloqueio para aplicação do seu método. Assim, para cada valor de limiar de cada rede são realizadas 1000 simulações e calculada a média para cada valor final. Pode-se observar no código abaixo uma representação do algoritmo de cálculo dos limiares 90% aplicado:

```
for(contador = 0; contador < 1000; contador++){
    while(maior componente conectado/componente original > 0,9){
        remover n de acordo com a estrategia de ataque
        re-calcular largura do maior componente conectado
    }

    resultados += qtd de n s restante/qtd de n s original
}

return 1-(soma resultados /1000);
```

### 4.4.2 Estratégias de ataque

Neste trabalho são utilizadas duas estratégias para simulação de falhas diferentes.

- Remoção aleatória: Simulação de falhas aleatórias em que o nó a ser removido é escolhido aleatoriamente e cada nó possui a mesma probabilidade de ser removido.
- Remoção com base no grau: Simulação de ataque direcionados, realizada através de remoção de nós com base no grau do mesmo. O nó a ser removido possui o maior grau, se dois ou mais nós possuem o mesmo grau, eles possuem a mesma probabilidade de ser removido. O maior grau é determinado utilizando a matriz Laplaciana que representa o grafo.

### 4.4.3 Redes Neurais Artificiais

A rede neural aplicada nesta pesquisa consiste de um *multilayer perceptron* que contém na camada de entrada um total de 6 nós, em que 4 dos nós da entrada recebem inicialmente os valores de um subconjunto do total de métricas que foram selecionadas a partir da análise de correlação que identifica as métricas mais importantes para representar sua respectiva rede. Os outros 2 nós restantes recebem inicialmente os valores da quantidade total de nós e de *links* de sua respectiva rede original.

Ainda, a rede possui apenas uma camada escondida que contém o uma quantidade de nós igual a 13 nós seguindo a função  $(2x + 1)$  se  $x$  for a quantidade de nós da camada de entrada. Finalmente, na sua camada de saída a rede possui apenas um nó que representa o valor da métrica limiar desejada.

### 4.4.4 Algoritmo de Treinamento

Previamente à esta etapa foram determinadas estruturas de dados nomeadas padrões de treinamento que recebem os valores conhecidos das métricas e suas respectivas saídas na forma dos limiares. Após carregar todos os valores necessários dentro dos padrões, se inicia a etapa de treinamento. Além disso, no trecho de código a seguir, pode-se observar a aplicação dos conceitos de validação cruzada para limitar os problemas de *overfitting* e *backpropagation* para otimização do treinamento da rede neural. O *overfitting* é um termo estatístico que descreve a presença de uma espécie de ruído que pode ser confundido como uma relação entre as variáveis do cenário. A aplicação de validação cruzada para mitigar este ruído é uma forma de otimizar os resultados obtidos.

Parâmetros de treinamento:

```
public void treinamento(){
    for(iteracoes = 0; iteracoes < 200000; iteracoes++){
        erroTreinamento = 0;
        for (int j = 0; j < padroesTreinamento.size(); j++) {
            padrao = padroesTreinamento.get(ordem);
            processarEntrada(padrao);
            processarCamadaEscondida();
            erroIndividual = processarSaida(padrao);
            if (skipOutliers && erroIndividual > 1.5) {
                continue;
            }
            erroTreinamento += erroIndividual;
            realimentarSaida(taxaAprendizado, padrao);
        }
    }
}
```

```

        realimentarCamadaEscondida(taxaAprendizado , padrao);
    }
    erroTreinamento /= padroesTreinamento.size();
    errosTreinamento.add(erroTreinamento);
}
}

```

Posteriormente à etapa de treinamento, os pesos obtidos são salvos em um arquivo para posterior uso em próximas etapas.

#### 4.4.5 Algoritmo de Teste

Os testes são realizados iterativamente em conjunto com o treinamento. Previamente à esta etapa foram criadas estruturas de dados nomeadas padrões de teste que possuem o mesmo papel dos padrões de treinamento, que é a organização e o armazenamento dos valores das entradas e suas respectivas saídas na forma dos limiares. Em seguida, se inicia a etapa de teste que segue o seguinte algoritmo:

```

erroValidacao = 0;
for (PadraoTeste pt : padroesTeste) {
    Validacao = obterSaida(pt.getEntrada());
    erroValidacao +=
        (Erro Quadratico Medio)
        .calcular(pt.getSaida()[0], Validacao);
}
erroValidacao /= padroesTeste.size();
errosValidacao.add(erroValidacao);

```

## 4.5 Dataset

Para as etapas de treinamento e teste foi selecionado um conjunto contendo 100 redes reais em (KNIGHT et al., 2011). Para aplicação do método e obtenção dos resultados, foi escolhido um conjunto completamente diferente que contém 20 redes, algumas estão ilustradas na Figura 4, em que pode se observar a variedade das redes contidas no *dataset*. A rede Garr ilustrada na Figura 4d, por exemplo, é a rede nacional italiana composta por universidades voltadas para pesquisa. Na versão ilustrada, ela possui 61 nós e 89 *links*. Ela se espalha por quase todo o território italiano, em sua topologia é possível identificar visualmente a presença de alguns *hubs*. Também, entre as redes ilustradas na Fig. 4 está a GTS Slovakia (Fig. 4a) que contém 66 nós e 37 *links* que visualmente apresenta uma grande dependencia em 3 *hubs* principais. Este *dataset*

também contém algumas redes menores como a Grena Georgia (Fig. 4c) que contém poucos nós e de nenhuma a baixíssimas quantidades de redundâncias.

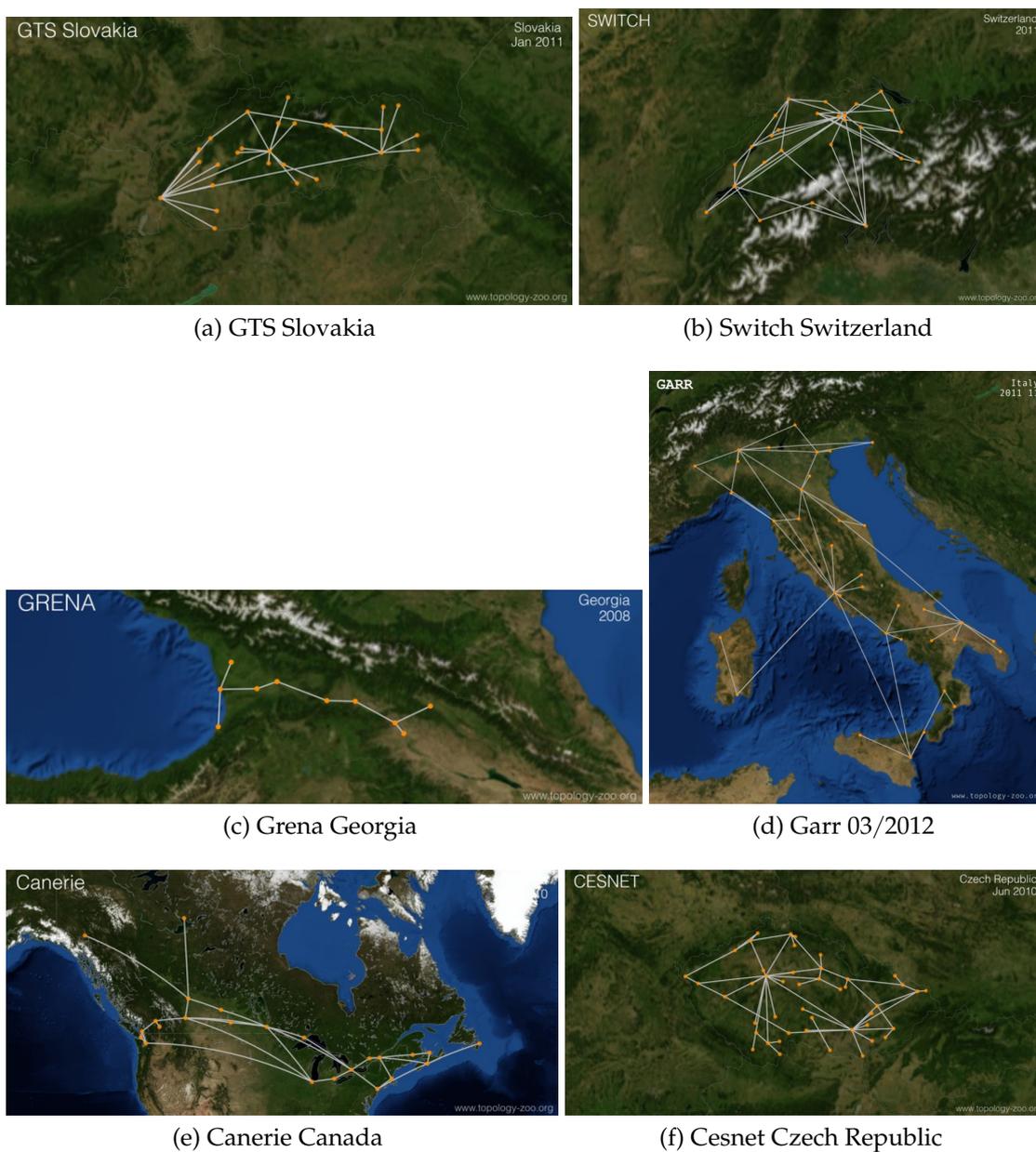


Figura 4 – Ilustração de algumas redes reais submetidas ao método proposto. Fonte: (KNIGHT et al., 2011)

## 5 Resultados

Este capítulo apresenta os resultados obtidos na implementação de cada etapa da pesquisa. Para provar a hipótese da pesquisa, de que é possível aplicar regressão numérica combinando métricas topológicas com baixo custo para obter um indicador de robustez de alto custo, é significativo que as margens do erro de treinamento, validação e aplicação da rede neural sejam pequenas. Para servir como parâmetro de comparação para análise dos valores do EQM obtidos nos resultados, será feita uma comparação com de Araújo et al (2015) em que é aplicada uma RNA treinada para realizar regressão numérica para produzir uma estimativa da probabilidade de bloqueio em redes ópticas comutadas por comprimento de onda.

### 5.1 Seleção de Métricas

A etapa de seleção de métricas tem como finalidade a obtenção de um subconjunto de métricas para serem utilizadas como entrada da RNA. O critério de seleção das métricas é que elas avaliem da melhor forma possível os diferentes aspectos da rede óptica estudada. Para tal, deseja-se aplicar um conjunto de métricas que avaliem diferentes aspectos das redes de computadores. Para quantificar a diversidade de aspectos estudados, utilizou-se o coeficiente produto-momento de Pearson. O coeficiente de Pearson apresenta valores +1 e -1 para conjuntos que apresentam forte correlações, ou seja, métricas que possuem altos valores de correlação estudam características semelhantes de uma rede. Assim, para que as métricas avaliadas possuam a maior diversidade de aspectos estudados possíveis, as métricas escolhidas apresentam os menores valores do coeficiente de correlação de Pearson.

Para esta etapa foram consideradas 7 métricas, que estão dispostas na Tabela 1.

Tabela 1 – Métricas topológicas aplicadas na pesquisa e suas representações.

Representação	Métrica
M1	Indicador de Robustez
M2	Condutância
M3	Conectividade Algébrica
M4	Assortatividade
M5	Entropia do grau de nós
M6	<i>DFT</i> da Entropia Laplaciana
M7	Conectividade Natural

Na Tabela 2, está representada a matriz de correlação entre as métricas topológicas da Tabela 1 em que estão contidos os valores do coeficiente produto-momento de Pearson calculado entre conjuntos que contém valores das métricas topológicas aplicadas a 100 redes do *dataset Topology-zoo*.

Tabela 2 – Matriz de correlação das métricas topológicas da Tabela 1.

	M1	M2	M3	M4	M5	M6	M7
M1	0	0,3803	0,7850	0,2767	0,0915	-0,0938	0,2242
M2	0,3803	0	0,7054	0,2872	-0,4391	-0,4308	-0,1770
M3	0,7850	0,7054	0	0,2726	-0,3999	-0,3403	0,0783
M4	0,2767	0,2872	0,2726	0	-0,0699	0,2374	-0,2356
M5	0,0915	-0,4391	-0,3999	-0,0699	0	0,3843	0,1452
M6	-0,0938	-0,4308	-0,3403	0,2374	0,3843	0	0,1290
M7	0,2242	-0,1770	0,0783	-0,2356	0,1452	0,1290	0

De Araújo et al (2015) apresenta um estudo que mostra o comportamento da média do Erro Quadrático Médio  $\overline{EQM}$  a medida que são inseridas variáveis na camada de entrada da RNA. Em seu estudo, de Araújo mostra que ocorre uma estabilização dos valores do  $\overline{EQM}$  a partir 5 variáveis na camada de entrada com o valor de  $2,83 \cdot 10^{-4}$ , assim como em 6 variáveis que apresenta valor de  $2,66 \cdot 10^{-4}$  e 7 variáveis que apresenta valor de  $2,57 \cdot 10^{-4}$ , quando comparados a um número menor de variáveis na camada de entrada.

Assim, nesta pesquisa se adotou a quantidade de 6 elementos na camada de entrada da RNA, em que 2 deles estão reservados para a quantidade de nós e de *links* da rede óptica estudada. Permitindo assim que 4 métricas sejam selecionadas para compor a camada de entrada da RNA.

Então, com o intuito da formação de subconjuntos contendo 4 métricas topológicas, decidiu-se a criar 7 subconjuntos de possíveis entradas para a RNA, em que cada uma das métricas da Tabela 1 atuaria como um núcleo para cada subconjunto. E tendo uma destas métricas topológicas como ponto de partida para cada subconjunto, selecionadas as métricas com os valores do coeficiente de Pearson mais próximos a 0 (zero). Desta maneira, analisando a Tabela 2 que contém a matriz total de correlação, foram formados os subconjuntos presentes na Tabela 3 que além da métrica núcleo, contém as 3 métricas com os valores de correlação mais próximos a 0 (zero) em relação ao núcleo do subconjunto.

Dos subconjuntos presentes na Tabela 3, foram realizados experimentos com todos os subconjuntos formados. O subconjunto de métricas topológicas que apresentou os menores valores de EQM nas etapas de treinamento e validação foi selecionado como parte da camada de entrada. Aquele que apresentou os menores erros de treinamento e

Tabela 3 – Subconjuntos de métricas topológicas formados após a análise da matriz de correlação. Ver Tabela 2

Núcleo	Métrica melhor correlacionada 1	Métrica melhor correlacionada 2	Métrica melhor correlacionada 3
M1	M5 (0,0915)	M6 (-0,0938)	M7 (0,2242)
M2	M7 (-0,1770)	M4 (0,2872)	M1 (0,3803)
M3	M7 (0,0783)	M4 (0,2726)	M6 (-0,3403)
M4	M5 (-0,0699)	M7 (-0,2356)	M6 (0,2374)
M5	M4 (-0,0699)	M1 (0,0915)	M7 (0,1452)
M6	M1 (-0,0938)	M7 (0,1290)	M4 (0,2374)
M7	M3 (0,0783)	M6 (0,1290)	M5 (0,1452)

validação foi o subconjunto que contém o indicador de robustez como núcleo.

Assim, o processo de seleção de métricas, resultou na escolha das métricas indicador de robustez como núcleo, entropia do grau de nós com coeficiente de 0,0915, DFT da entropia laplaciana com -0,0938 e conectividade natural com valor de 0,2242 de correlação em relação ao núcleo para abastecer a entrada da rede neural para ambas as previsões dos valores,  $f_{90\%}$  aleatória e direcionada, juntamente com as quantidades de nós e de *links*.

## 5.2 Treinamento/Validação da RNA

Esta etapa tem como finalidade a obtenção de uma RNA treinada com baixos valores de EQM que seja capaz de realizar uma previsão confiável dos valores do limiar  $f_{90\%}$  aleatório e direcionado. Para melhor apresentar os resultados nesta etapa, de Araújo et al (2015) aplica uma técnica de estatística descritiva que tem a finalidade de fornecer um resumo simples dos resultados, o diagrama de caixa (box-plot). Nele, é feita a representação gráfica que representa a variação dos valores do EQM no decorrer do treinamento e da validação das RNAs. Analisando-o, é possível determinar os estão 50% dos valores mais prováveis, a mediana e os valores extremos do treinamento e da validação do limiar estudado. Ele contém o fio (*whisker*) que se estendem verticalmente ou horizontalmente indicando as variações fora da caixa. Ainda, o diagrama de caixa apresenta o elemento de valores discrepantes (*outliers*) que podem ser plotados individualmente.

A Fig. 5 contém o gráfico *box-plot* do Erro Quadrático Médio relacionado às fases de treinamento e validação das redes utilizadas para previsão de ambos limiares, comparando com os resultados obtidos em de Araújo et al (2015), ele indica que a rede é adequada para a execução das previsões desejadas já que ele afirma a obtenção de uma margem de erro reduzida. Em geral, na Fig. 5 cada caixa corresponde à

estatística descritiva de 30 execuções da aplicação das redes neurais, que correspondem ao tamanho do conjunto de validação, inicializadas aleatoriamente, que fornece uma visão geral sobre as margens de erro obtidas. A média do EQM para o conjunto de validação foi de  $1,19 \cdot 10^{-4}$  para falhas direcionadas e  $1,12 \cdot 10^{-3}$  para falhas aleatórias.

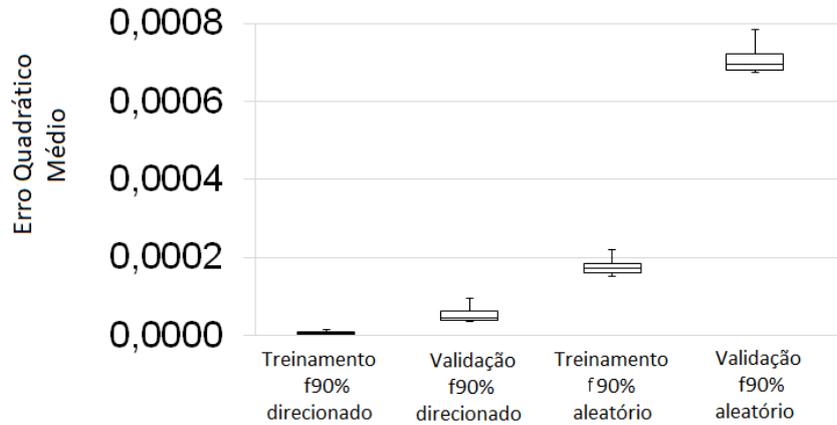
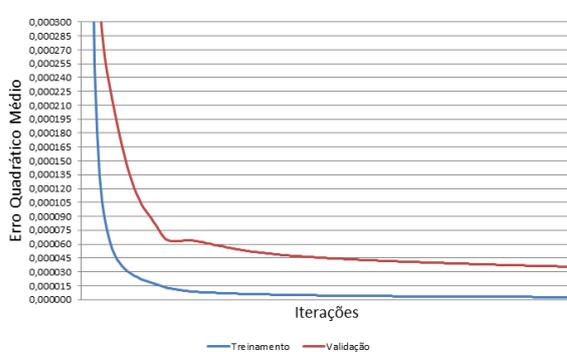


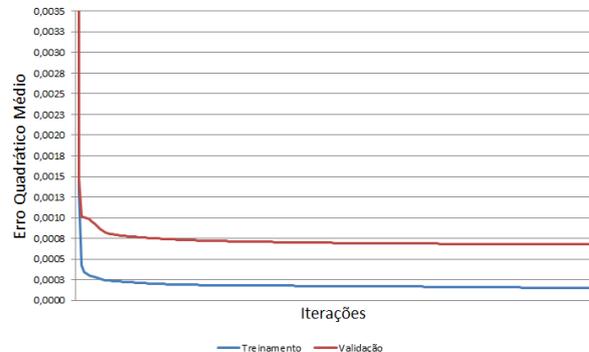
Figura 5 –  $\overline{EQM}$  dos limiares  $f_{90}\%$  para o modelo escolhido para ambas simulações de falhas (direcionada e aleatória). Fonte: O autor.

Os valores dos *outliers* máximos do diagrama de caixa da Figura 5 estão presentes fora da área de plotagem por apresentarem valores na grandeza de  $10^{-1}$  o que dificultaria a visibilidade dos demais valores. Estes valores *outliers* são gerados nas iterações iniciais da execução das fase de treinamento e validação para ambos os limiares, como podemos identificar nas Figuras 6b e 6a.

As Figuras 6a e 6b representam o comportamento do EQM no decorrer das 200.000 iterações do treinamento/validação da RNA. Nelas observa-se uma estabilização das curvas de erro com tendência ao valor 0 (zero) após por volta das 2000 iterações iniciais para o limiar direcionado e antes das 500 iterações iniciais para o limiar aleatório.



(a) Curva do limiar direcionado.



(b) Curva do limiar aleatório.

Figura 6 – Desenvolvimento do EQM no decorrer das iterações. Fonte: O autor.

### 5.3 Aplicação da RNA treinada

Tendo em vista que na etapa de treinamento e validação da RNA, os valores do EQM foram baixos, quando verificados conjuntamente aos valores obtidos por de Araújo et al (2015), nesta etapa espera-se que os valores obtidos pela aplicação da RNA para predição dos valores de limiares demonstre que as margens de erro geradas sejam tão baixas quanto a etapa de treinamento e validação.

Analisando algumas das redes ilustradas na Fig. 4 e seus respectivos valores nas Tabelas 4 e 5, como a rede GTS Slovakia (Fig. 4a), percebe-se que para simulação de falhas direcionadas é necessário que sejam removidos em torno de 3,5% da quantidade de nós original da rede e o valor do erro quadrático médio da diferença entre os valores obtidos por simulação e pelo método proposto através da RNA é de  $3,10 \cdot 10^{-7}$ , já para os valores do limiar  $f_{90\%}$  aleatório, mostra-se que em o valor da simulação aponta que 9,1% dos nós originais da rede devem ser removidos, enquanto a RNA determina que aproximadamente 8,2% dos nós originais devem ser removidos o que fornece uma maior margem de erro igual a  $8,89 \cdot 10^{-5}$ . Para o limiar aleatório disposto na Tabela 5, a aplicação que obteve a menor margem de erro foi a rede *Garr Italy 03/2012* que apresenta o EQM de  $2,41 \cdot 10^{-6}$  já a pior margem de erro foi a rede *Digex US* que apresenta um EQM de  $1,32 \cdot 10^{-3}$ . Para o limiar direcionado, a aplicação com menor margem de erro foi a *NSF US* que apresenta um valor do EQM de  $3,64 \cdot 10^{-9}$  e a que apresenta maior margem de erro foi a *FatMan UK* com  $7,31 \cdot 10^{-4}$  de EQM.

Assim, observa-se que a abordagem de falhas aleatórias apresenta um desafio maior para a rede neural. Entretanto, a característica mais relevante desta abordagem é a diferença entre os valores do tempo computacional necessários para a estimativa dos valores dos limiares. A simulação de falhas dura em torno de 13466 ms para estimar o valor do limiar  $f_{90\%}$ . Enquanto a RNA necessita em torno de 45 ms para realizar a mesma operação. Assim observa-se uma aceleração em 300 vezes do processo.

As Tabelas 5 e 4 apresentam os resultados obtidos a partir da aplicação do método proposto em comparação com os valores dos limiares aleatórios e direcionados obtidos a partir da execução das simulações. Em geral, os resultados obtidos foram expressivos utilizando um *dataset* pequeno que contém apenas 100 redes reais. A partir dos resultados obtidos, acredita-se que se torna possível uma diminuição drástica nos recursos que são alocados para realização de simulações direcionadas à quantificação de robustez em sistemas. Além disso, como existe a possibilidade da utilização deste método em termos de notação de grafos, torna-se possível a aplicação deste modelo em problemas similares em outras área de aplicação, como nas redes sociais ou redes ecológicas, em que são direcionadas grandes quantidades de recursos para realização de simulações. Outro fator é a diminuição de recursos temporais, que até mesmo para área de pesquisas científicas é de grande ajuda já que a diminuição do investimento de

Tabela 4 – Métricas de robustez e suas respectivas margens de erro 20 redes ópticas selecionadas do *dataset* (KNIGHT et al., 2011) para limiares  $f_{90\%}$  direcionado. Fonte: O autor.

Rede	$f_{90\%}$ para falhas direcionadas	RNA $f_{90\%}$ para falhas direcionadas	Erro Quadrático para falhas direcionadas
<i>GTS Slovakia</i>	0,0357	0,0352	$3,10 \cdot 10^{-7}$
<i>Switch Switzerland</i>	0,0333	0,0317	$2,51 \cdot 10^{-6}$
<i>Grena Georgia</i>	0,0769	0,0824	$2,97 \cdot 10^{-5}$
<i>Garr Italy 03/2012</i>	0,0222	0,0221	$2,78 \cdot 10^{-8}$
<i>Canerie Canada</i>	0,0417	0,0430	$1,88 \cdot 10^{-6}$
<i>Cesnet Czech Republic</i>	0,0294	0,0287	$4,97 \cdot 10^{-7}$
<i>Belnet Belgium</i>	0,0476	0,0423	$2,85 \cdot 10^{-5}$
<i>Arpanet US</i>	0,1111	0,1107	$1,47 \cdot 10^{-7}$
<i>Highwinds</i>	0,0556	0,0617	$3,81 \cdot 10^{-5}$
<i>Garr Italy 12/2001</i>	0,0454	0,0443	$1,39 \cdot 10^{-6}$
<i>RHnet</i>	0,0769	0,0783	$2,02 \cdot 10^{-6}$
<i>Noel Communications US</i>	0,0526	0,0540	$1,96 \cdot 10^{-6}$
<i>GTS Poland</i>	0,0385	0,0368	$2,64 \cdot 10^{-6}$
<i>Digex US</i>	0,0323	0,0302	$4,15 \cdot 10^{-6}$
<i>Garr Italy 01/2011</i>	0,0227	0,0224	$1,10 \cdot 10^{-7}$
<i>Packet Exchange</i>	0,0476	0,0479	$5,94 \cdot 10^{-8}$
<i>IBM US</i>	0,0556	0,0554	$1,50 \cdot 10^{-8}$
<i>Sanet Slovakia</i>	0,0303	0,0296	$4,95 \cdot 10^{-7}$
<i>NSF US</i>	0,0769	0,0770	$3,64 \cdot 10^{-9}$
<i>FatMan UK</i>	0,25	0,2230	$7,31 \cdot 10^{-4}$

tempo em simulações abre espaço para realização de novas atividades com intuito de enriquecer esta área de desenvolvimento.

Tabela 5 – Métricas de robustez e suas respectivas margens de erro 20 redes ópticas selecionadas do *dataset* (KNIGHT et al., 2011) para limiares  $f_{90\%}$  aleatório.

<b>Rede</b>	<b><math>f_{90\%}</math> para falhas aleatórias</b>	<b>RNA <math>f_{90\%}</math> para falhas aleatórias</b>	<b>Erro Quadrático para falhas aleatórias</b>
<i>GTS Slovakia</i>	0,0912	0,0818	$8,89 \cdot 10^{-5}$
<i>Switch Switzerland</i>	0,0955	0,0931	$5,86 \cdot 10^{-6}$
<i>Grena Georgia</i>	0,1069	0,0888	$3,27 \cdot 10^{-4}$
<i>Garr Italy 03/2012</i>	0,0813	0,0828	$2,41 \cdot 10^{-6}$
<i>Canerie Canada</i>	0,1094	0,0994	$9,90 \cdot 10^{-5}$
<i>Cesnet Czech Republic</i>	0,0811	0,0738	$5,21 \cdot 10^{-5}$
<i>Belnet Belgium</i>	0,1300	0,1086	$4,58 \cdot 10^{-4}$
<i>Arpanet US</i>	0,1111	0,1242	$1,72 \cdot 10^{-4}$
<i>Highwinds</i>	0,1054	0,0988	$4,47 \cdot 10^{-5}$
<i>Garr Italy 12/2001</i>	0,1128	0,0950	$3,16 \cdot 10^{-4}$
<i>RHnet</i>	0,1309	0,0993	$9,98 \cdot 10^{-4}$
<i>Noel Communications US</i>	0,0968	0,1056	$7,78 \cdot 10^{-5}$
<i>GTS Poland</i>	0,0815	0,0833	$3,40 \cdot 10^{-6}$
<i>Digex US</i>	0,1099	0,0736	$1,32 \cdot 10^{-3}$
<i>Garr Italy 01/2011</i>	0,0828	0,0844	$2,50 \cdot 10^{-6}$
<i>Packet Exchange</i>	0,1194	0,0996	$3,91 \cdot 10^{-4}$
<i>IBM US</i>	0,1082	0,1064	$3,22 \cdot 10^{-6}$
<i>Sanet Slovakia</i>	0,0888	0,0532	$1,27 \cdot 10^{-3}$
<i>NSF US</i>	0,1339	0,1126	$4,56 \cdot 10^{-4}$
<i>FatMan UK</i>	0,25	0,2735	$5,54 \cdot 10^{-4}$

## 6 Conclusões

Neste trabalho, foi avaliada a possibilidade da obtenção de medidas de robustez topológica que possuem alto custo computacional e temporal a partir de um método alternativo que utiliza métricas correlacionadas simples (com baixo custo) como entrada para um método de aprendizado de máquina supervisionado usando um modelo de rede neural *multilayer perceptron* para prever valores de métricas custosas através de regressão.

Na etapa de treinamento e validação da RNA, analisando os valores obtidos, através do cálculo do erro quadrático médio, pode-se fazer uma comparação com resultados de trabalhos similares que também aplicam RNA para regressão. Como em de Araújo et al que aplica regressão para determinar probabilidade de bloqueio, que é uma medida de desempenho, em redes ópticas e chegar à conclusão que os valores de erro obtidos satisfazem as necessidades de uma solução viável, ou seja, as margens de erro obtidas são baixas o suficiente para predizer os valores desejados com proximidade suficiente para aplicações práticas. E para analisar esta etapa, a Figura 5 que ilustra o gráfico *box-plot* também fornece uma visão geral das margens de erro obtidas que reforça as conclusões extraídas.

Na etapa de aplicação, as métricas topológicas previamente selecionadas foram utilizadas como entrada da RNA e após o processamento, resultaram em valores de limiares  $f_{90\%}$  para falhas direcionadas e aleatórias e também obtiveram uma margem erro satisfatória. Finalmente, analisando os resultados, pode-se inferir que é possível realizar a previsão dos limiares de robustez com erros entre  $10^{-3}$  e  $10^{-9}$  com uma velocidade de até 300 vezes mais rápida.

### 6.1 Artigos Publicados

Os resultados desta pesquisa foram sumarizados em um artigo científico que foi aceito para publicação no evento internacional IEEE 2017 LA-CCI.

### 6.2 Trabalhos Futuros

Como trabalhos futuros, pode-se analisar a viabilidade da aplicação de um método similar que seleciona métricas com base no menor grau de correlação entre as mesma. Ainda, pode-se analisar outras técnicas de aprendizado de máquina para regressão. Além disso, é possível analisar o desempenho das redes neurais para a

predição de robustez de redes complexas utilizando um *dataset* escalável.

## Referências

ALENAZI MOHAMMED JF; STERBENZ, J. P. Evaluation and comparison of several graph robustness metrics to improve network resilience. *Reliable Networks Design and Modeling (RNDM), 7th International Workshop on. IEEE, 2015. p. 7-13.*, 2015. Citado 3 vezes nas páginas [11](#), [14](#) e [25](#).

ARAUJO DANILO RB; BASTOS FILHO, C. J. M. F. J. F. Artificial neural networks to estimate blocking probability of transparent optical networks: A robustness study for different networks. *17th International Conference on Transparent Optical Networks (ICTON). IEEE. p. 1-4.*, 2015. Citado 2 vezes nas páginas [22](#) e [26](#).

ARAUJO DANILO RB; BASTOS FILHO, C. J. M. F. J. F. D. Methodology to obtain a fast and accurate estimator for blocking probability of optical networks. *Journal of Optical Communications and Networking, v. 7, n. 5, p. 380-391*, 2015. Citado 2 vezes nas páginas [25](#) e [26](#).

ARAUJO DANILO RB; MARTINS FILHO, J. F. B. F. C. J. D. Non-supervised learning applied to analysis of topological metrics of optical networks. *Designing with Computational Intelligence. Springer International Publishing. p. 109-126.*, 2017. Citado na página [26](#).

ARAÚJO DANILO RB; BASTOS-FILHO, C. J. M.-F. J. F. D. Using the entropy of the dft of the laplacian eigenvalues to assess networks. *Complex Networks V. Springer, Cham. p. 209-216*, 2014. Citado na página [20](#).

BISHOP, C. *Pattern Recognition and Machine Learning (Information Science and Statistics), 2nd printing edn.* [S.l.: s.n.], 2007. Citado 2 vezes nas páginas [13](#) e [18](#).

CHANG LIANG; WU, Z. Performance and reliability of electrical power grids under cascading failures. *International Journal of Electrical Power and Energy Systems, v. 33, n. 8, p. 1410-1419.*, 2011. Citado 2 vezes nas páginas [11](#) e [12](#).

DEMETRIUS LLOYD; MANKE, T. Robustness and network evolution—an entropic principle. *Physica A: Statistical Mechanics and its Applications, v. 346, n. 3, p. 682-696*, 2005. Citado na página [20](#).

HAYKIN, S. S. *Neural networks and learning machines.* [S.l.: s.n.], 2010. Citado na página [22](#).

JAMAKOVIC A.; UHLIG, S. On the relationship between the algebraic connectivity and graph's robustness to node and link failures. *Next Generation Internet Networks, 3rd EuroNGI Conference on IEEE, p. 96-102.*, 2007. Citado 4 vezes nas páginas [11](#), [13](#), [14](#) e [25](#).

JAMAKOVIC A.; VAN MIEGHEM, P. On the robustness of complex networks by using the algebraic connectivity. *International Conference on Research in Networking. Springer Berlin Heidelberg. p. 183-194.*, 2008. Citado 4 vezes nas páginas [12](#), [13](#), [17](#) e [19](#).

KNIGHT, J. C. Safety critical systems: challenges and directions. *Software Engineering, 2002. ICSE 2002. Proceedings of the 24rd International Conference on IEEE.*, 2002. Citado 2 vezes nas páginas [11](#) e [16](#).

- KNIGHT, S. et al. The internet topology zoo. *IEEE Journal on Selected Areas in Communications*, IEEE, v. 29, n. 9, p. 1765–1775, 2011. Citado 7 vezes nas páginas 3, 4, 15, 32, 33, 39 e 40.
- LEUNG, C.; CHAU, H. Weighted assortative and disassortative networks model. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 378, n. 2, p. 591–602, 2007. Citado na página 18.
- LEWIS, T. G. Network science: Theory and applications. Wiley Publishing, 2009. Citado na página 18.
- LUO, X. S. e. a. Analysis of cascading failure in complex power networks under the load local preferential redistribution rule. *Physica A: Statistical Mechanics and its Applications*, v. 391, n. 8, p. 2771–2777, 2012. Citado na página 11.
- PAI, G. V. Genetic algorithm based backpropagation networks. *NEURAL NETWORKS, FUZZY SYSTEMS AND EVOLUTIONARY ALGORITHMS: SYNTHESIS AND APPLICATIONS*, PHI Learning Pvt. Ltd., p. 399, 2017. Citado na página 24.
- QIU, T. e. a. A greedy model with small world for improving the robustness of heterogeneous internet of things. *Computer Networks*, v. 101, p. 127–143., 2016. Citado 2 vezes nas páginas 27 e 28.
- RAMASWAMI RAJIV; SIVARAJAN, K. S.-G. Optical networks: a practical perspective. *Morgan Kaufmann*, 2009. Citado 3 vezes nas páginas 11, 14 e 17.
- SARKAR TUHIN; ROOZBEHANI, M. D.-M. A. Robustness scaling in large networks. *American Control Conference (ACC). IEEE*. p. 197–202., 2016. Citado 2 vezes nas páginas 26 e 27.
- TANG, L. e. a. Robustness of assembly supply chain networks by considering risk propagation and cascading failure. *Physica A: Statistical Mechanics and its Applications*, v. 459, p. 129–139., 2016. Citado na página 27.
- TRAN, H. T.; DOMERÇANT, J. C.; MAVRIS, D. N. A network-based cost comparison of resilient and robust system-of-systems. *Procedia Computer Science*, Elsevier, v. 95, p. 126–133, 2016. Citado na página 16.
- WANG, X. e. a. Quantifying the robustness of metro networks. *arXiv preprint arXiv:1505.06664*, 2015. Citado 6 vezes nas páginas 11, 12, 14, 18, 20 e 25.