UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Classificação de sinais de epilepsia utilizando redes complexas

Daniel Moreira Cestari

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Daniel Moreira Cestari

Classificação de sinais de epilepsia utilizando redes complexas

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientadora: Prof. Dr. João Luís Garcia Rosa

USP – São Carlos
Julho de 2017

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

C421c Cestari, Daniel Moreira
Classificação de sinais de epilepsia utilizando
redes complexas / Daniel Moreira Cestari;
orientador João Luis Garcia Rosa. -- São Carlos,
2017.
192 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2017.

1. Classificação. 2. Redes complexas. 3. EEG. 4.
Epilepsia. I. Rosa, João Luis Garcia, orient. II.
Título.

Daniel Moreira Cestari

Classification of epileptic signals using complex networks

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. João Luís Garcia Rosa

USP – São Carlos
July 2017

AGRADECIMENTOS

Agradeço a minha família pela compreensão e frequente privação de convívio.

Aos amigos do laboratório por aguentar todas as brincadeiras.

Ao CNPq pelo suporte financeiro.

Ao CEPID-CeMEAI pelo recurso computacional disponibilizado.

Por fim mas não menos importante, agradecimentos especiais às burocracias que têm único propósito testar nossa paciência.

RESUMO

CESTARI, D. M. **Classificação de sinais de epilepsia utilizando redes complexas**. 2017. 192 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

Contexto: Epilepsia não é uma única doença, mas uma família de síndromes que compartilham a recorrência de crises. Estima-se que 3% da população em geral terá epilepsia em algum momento em suas vidas. A detecção de crises epiléticas é frequentemente feita através da análise de exames de eletroencefalografia. Há várias dificuldades na detecção de crises, variabilidade entre pessoas, localização do conteúdo espectral, interferências, dentre outras. Motivação: Há um crescente uso com bons resultados de redes complexas para análise de séries temporais, mas poucos destes são voltados à análise de sinais de epilepsia. Os trabalhos que analisam epilepsia, em geral, negligenciam uma análise estatística rigorosa. Ainda há dúvida quanto à utilização de algoritmos prospectivos para predição de crises. Métodos: As séries temporais são analisadas utilizando 7 tamanhos diferentes de janelas, 256, 303, 512, 910, 1.024, 2.048, e 2.730 pontos. São utilizados 6 algoritmos de conversão de série temporal em rede complexa, redes de k vizinhos mais próximos, redes de k vizinhos mais próximos adaptativos, redes de *epsilon* vizinhança, redes cíclicas, redes de transição, e grafos de visibilidade. Cada um desses algoritmos têm seus parâmetros, e no total são realizadas 75 conversões. Para cada rede complexa gerada, são extraídas 21 medidas que as caracterizam. Com a extração dessas medidas, um novo conjunto de dados é formado e utilizado para treinar 37 classificadores diferentes, divididos em 4 classes, análise de discriminante linear, árvore de decisão, k vizinhos mais próximos, e máquina de vetores de suporte. É utilizada uma validação cruzada com 10-*folds* numa parte do conjunto de dados separada para o treino dos classificadores, e apenas o melhor classificador dentre os 37 foi selecionado em cada conversão realizada. No conjunto de teste, é feita a estimativa de desempenho do melhor classificador, que é então comparado à um preditor aleatório e ao estado da arte. Resultados: A rede de *epsilon* vizinhança obteve o melhor resultado, com 100% de acurácia no conjunto de teste em quase todos os cenários, com janelas de tamanho pequeno e com a análise de discriminante linear. As outras redes também tiveram bons resultados, comparáveis ao estado da arte, exceto a rede de transição cujo desempenho foi ruim. Conclusão: Foi possível desenvolver um algoritmo prospectivo com classificador linear utilizando a rede de *epsilon* vizinhança, com desempenho comparável ao estado da arte e com rigorosa avaliação estatística, e não apenas utilizando a acurácia como medida de desempenho.

Palavras-chave: Classificação, Redes Complexas, EEG, Epilepsia.

ABSTRACT

CESTARI, D. M. **Classification of epileptic signals using complex networks**. 2017. 192 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

Context: Epilepsy is not a single disease, but a family of syndromes that share recurrent seizures. It is estimated that 3% of the population will have epilepsy at some moment of their life. Seizure detection is frequently done through EEG analysis. There are several difficulties in seizure detection, people variability, the location of the spectral content, interferences, among other things. **Motivation:** There is a growing usage with good results of the complex networks to analyze time series, but few studies focusing on epilepsy. The works that have analyzed epilepsy, in general, have neglected a strict statistical analysis. There is still doubts regarding the usage of prospective algorithms to predict seizures. **Methods:** The time series were analyzed on 7 different window sizes, 256, 303, 512, 910, 1024, 2048, and 2730 points. We used 6 different algorithms to convert the time series into complex networks, k nearest neighbors network, adaptive k nearest neighbors network, epsilon neighborhood network, cycle network, transition network, visibility graph. Each algorithm has its parameters, and in total, we performed 75 conversions. For each conversion, the network extracted 21 measures. A new dataset is formed with these measures, and it was used to train 37 classifiers, divided into 4 classes, linear discriminant analysis, decision tree, k nearest neighbors, support vector machine. We used 10-fold cross-validation in a training set, separated from the whole dataset, and only the best classifier between the 37 was selected for each conversion. In the test set, we estimated the performance of the best classifiers, and then they were compared with a random predictor and with the state-of-the-art. **Results:** The epsilon neighborhood network presented the best result with 100% accuracy over almost all scenarios in the test set, with small window sizes and the linear discriminant analysis. The other networks also had good results, comparable to the state-of-the-art, except the transition network which had poor performance. **Conclusion:** We were able to develop a prospective algorithm with a linear classifier using the epsilon neighborhood network, with a performance comparable to the state-of-the-art and with rigorous statistical analysis, and not only using the accuracy as our performance measure.

Keywords: Classification, Complex Networks, EEG, Epilepsy.

LISTA DE ILUSTRAÇÕES

Figura 1 – Etapas utilizadas na classificação.	27
Figura 2 – Sistema 10-20 de posicionamento de eletrodos (WIKIPEDIA, 2016).	30
Figura 3 – Método para classificação de EEG. Adaptado de (ZHU; LI; WEN, 2014).	32
Figura 4 – Exemplo de um grafo de visibilidade. Adaptado de (LACASA <i>et al.</i> , 2008).	48
Figura 5 – Séries da Universidade de Bonn. A - Paciente saudável com olhos abertos. B - Sinal de EEG durante uma crise.	55
Figura 6 – Trecho do conjunto da série estocástica usada.	56
Figura 7 – Trecho do sinal determinístico (sistema de <i>Lorenz</i>) usado. A - Atrator do primeiro conjunto de parâmetros. B - Atrator do segundo conjunto de parâmetros. C - Série os dois conjuntos de parâmetros repetidos suas vezes.	57
Figura 8 – Exemplo de agrupamento	63
Figura 9 – Esquema geral de predição de epilepsia. Adaptado de (MORMANN, 2008).	65
Figura 10 – Resultados da análise determinística (window size and dynamic threshold).	76
Figura 11 – Resultados da análise determinística (epsilon and embedding dimension).	77
Figura 12 – Resultados da análise estocástica.	78
Figura 13 – Medidas de uma janela de uma série do conjunto F utilizando uma rede cíclica.	82
Figura 14 – Desempenho no conjunto de treino da rede <i>aknn</i> para as classes FS e NS.	85
Figura 15 – Desempenho no conjunto de treino da rede <i>aknn</i> para as classes NFS e ONF.	86
Figura 16 – Desempenho no conjunto de treino da rede <i>aknn</i> para as classes OS.	87
Figura 17 – Desempenho no conjunto de treino da rede <i>knn</i> para as classes FS e NS.	88
Figura 18 – Desempenho no conjunto de treino da rede <i>knn</i> para as classes NFS e ONF.	89
Figura 19 – Desempenho no conjunto de treino da rede <i>knn</i> para as classes OS.	90
Figura 20 – Desempenho no conjunto de treino da rede <i>eps</i> para as classes FS e NS.	91
Figura 21 – Desempenho no conjunto de treino da rede <i>eps</i> para as classes NFS e ONF.	92
Figura 22 – Desempenho no conjunto de treino da rede <i>eps</i> para as classes OS.	93
Figura 23 – Desempenho no conjunto de treino da rede <i>cycle</i> para as classes FS e NS.	94
Figura 24 – Desempenho no conjunto de treino da rede <i>cycle</i> para as classes NFS e ONF.	95
Figura 25 – Desempenho no conjunto de treino da rede <i>cycle</i> para as classes OS.	96
Figura 26 – Desempenho no conjunto de treino da rede <i>trans</i> para as classes FS e NS.	97
Figura 27 – Desempenho no conjunto de treino da rede <i>trans</i> para as classes NFS e ONF.	98
Figura 28 – Desempenho no conjunto de treino da rede <i>trans</i> para as classes OS.	99
Figura 29 – Desempenho no conjunto de treino da rede <i>vis</i> para as classes FS e NS.	100
Figura 30 – Desempenho no conjunto de treino da rede <i>vis</i> para as classes NFS e ONF.	101

Figura 31 – Desempenho no conjunto de treino da rede <i>vis</i> para as classes OS.	102
Figura 32 – Desempenho no conjunto de teste da rede <i>aknn</i> para as classes FS e NS. . .	104
Figura 33 – Desempenho no conjunto de teste da rede <i>aknn</i> para as classes NFS e ONF. .	105
Figura 34 – Desempenho no conjunto de teste da rede <i>aknn</i> para as classes OS.	106
Figura 35 – Desempenho no conjunto de teste da rede <i>knn</i> para as classes FS e NS. . . .	107
Figura 36 – Desempenho no conjunto de teste da rede <i>knn</i> para as classes NFS e ONF. .	108
Figura 37 – Desempenho no conjunto de teste da rede <i>knn</i> para as classes OS.	109
Figura 38 – Desempenho no conjunto de teste da rede <i>eps</i> para as classes FS e NS. . . .	110
Figura 39 – Desempenho no conjunto de teste da rede <i>eps</i> para as classes NFS e ONF. .	111
Figura 40 – Desempenho no conjunto de teste da rede <i>eps</i> para as classes OS.	112
Figura 41 – Desempenho no conjunto de teste da rede <i>cycle</i> para as classes FS e NS. . .	113
Figura 42 – Desempenho no conjunto de teste da rede <i>cycle</i> para as classes NFS e ONF.	114
Figura 43 – Desempenho no conjunto de teste da rede <i>cycle</i> para as classes OS.	115
Figura 44 – Desempenho no conjunto de teste da rede <i>trans</i> para as classes FS e NS. . .	116
Figura 45 – Desempenho no conjunto de teste da rede <i>trans</i> para as classes NFS e ONF.	117
Figura 46 – Desempenho no conjunto de teste da rede <i>trans</i> para as classes OS.	118
Figura 47 – Desempenho no conjunto de teste da rede <i>vis</i> para as classes FS e NS. . . .	119
Figura 48 – Desempenho no conjunto de teste da rede <i>vis</i> para as classes NFS e ONF. .	120
Figura 49 – Desempenho no conjunto de teste da rede <i>vis</i> para as classes OS.	121
Figura 50 – <i>Boxplot</i> do número de vértices da rede de vizinhos mais próximos adaptativos com parâmetro <i>E0</i> igual a 7.	148
Figura 51 – <i>Boxplot</i> do número de arestas da rede de vizinhos mais próximos adaptativos com parâmetro <i>E0</i> igual a 7.	149
Figura 52 – <i>Boxplot</i> do número de vértices da rede de vizinhos mais próximos adaptativos com parâmetro <i>E0</i> igual a 13.	150
Figura 53 – <i>Boxplot</i> do número de arestas da rede de vizinhos mais próximos adaptativos com parâmetro <i>E0</i> igual a 13.	151
Figura 54 – <i>Boxplot</i> do número de vértices da rede de vizinhos mais próximos adaptativos com parâmetro <i>E0</i> igual a 14.	152
Figura 55 – <i>Boxplot</i> do número de arestas da rede de vizinhos mais próximos adaptativos com parâmetro <i>E0</i> igual a 14.	153
Figura 56 – <i>Boxplot</i> do número de vértices da rede de vizinhos mais próximos adaptativos com parâmetro <i>E0</i> igual a 17.	154
Figura 57 – <i>Boxplot</i> do número de arestas da rede de vizinhos mais próximos adaptativos com parâmetro <i>E0</i> igual a 17.	155
Figura 58 – <i>Boxplot</i> do número de vértices da rede de vizinhos mais próximos com parâmetro <i>k</i> igual a 9.	156
Figura 59 – <i>Boxplot</i> do número de arestas da rede de vizinhos mais próximos com parâ- metro <i>k</i> igual a 9.	157

Figura 60 – <i>Boxplot</i> do número de vértices da rede de vizinhos mais próximos com parâmetro k igual a 12.	158
Figura 61 – <i>Boxplot</i> do número de arestas da rede de vizinhos mais próximos com parâmetro k igual a 12.	159
Figura 62 – <i>Boxplot</i> do número de vértices da rede de vizinhos mais próximos com parâmetro k igual a 13.	160
Figura 63 – <i>Boxplot</i> do número de arestas da rede de vizinhos mais próximos com parâmetro k igual a 13.	161
Figura 64 – <i>Boxplot</i> do número de vértices da rede de <i>epsilon</i> vizinhança com parâmetro <i>epsilon</i> igual a 1,2.	163
Figura 65 – <i>Boxplot</i> do número de arestas da rede de <i>epsilon</i> vizinhança com parâmetro <i>epsilon</i> igual a 1,2.	164
Figura 66 – <i>Boxplot</i> do número de vértices da rede de <i>epsilon</i> vizinhança com parâmetro <i>epsilon</i> igual a 1,7.	165
Figura 67 – <i>Boxplot</i> do número de arestas da rede de <i>epsilon</i> vizinhança com parâmetro <i>epsilon</i> igual a 1,7.	166
Figura 68 – <i>Boxplot</i> do número de vértices da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,01.	167
Figura 69 – <i>Boxplot</i> do número de arestas da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,01.	168
Figura 70 – <i>Boxplot</i> do número de vértices da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,1.	169
Figura 71 – <i>Boxplot</i> do número de arestas da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,1.	170
Figura 72 – <i>Boxplot</i> do número de vértices da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,2.	171
Figura 73 – <i>Boxplot</i> do número de arestas da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,2.	172
Figura 74 – <i>Boxplot</i> do número de vértices da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,5.	173
Figura 75 – <i>Boxplot</i> do número de arestas da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,5.	174
Figura 76 – <i>Boxplot</i> do número de vértices da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,7.	175
Figura 77 – <i>Boxplot</i> do número de arestas da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,7.	176
Figura 78 – <i>Boxplot</i> do número de vértices da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,9.	177

Figura 79 – <i>Boxplot</i> do número de arestas da rede cíclica com parâmetro <i>limiar de correlação</i> igual a 0,9.	178
Figura 80 – <i>Boxplot</i> do número de vértices do grafo de vizinhança.	179
Figura 81 – <i>Boxplot</i> do número de arestas do grafo de vizinhança.	180
Figura 82 – <i>Boxplot</i> da medidas da rede de vizinhos mais próximos adaptativos.	182
Figura 83 – <i>Boxplot</i> da medidas da rede de vizinhos mais próximos.	183
Figura 84 – <i>Boxplot</i> da medidas da rede de <i>epsilon</i> vizinhança.	184
Figura 85 – <i>Boxplot</i> da medidas da rede cíclica.	185
Figura 86 – <i>Boxplot</i> da medidas da rede de transição.	186
Figura 87 – <i>Boxplot</i> da medidas do grafo de visibilidade.	187
Figura 88 – Contagem de cada atributo após a seleção de atributos.	191

LISTA DE TABELAS

Tabela 1 – Tabela de contingência	63
Tabela 2 – Informações gerais das conversões	80
Tabela 3 – Melhores classificadores para classe FS	123
Tabela 4 – Melhores classificadores para classe NS	123
Tabela 5 – Melhores classificadores para classe NFS	124
Tabela 6 – Melhores classificadores para classe ONF	124
Tabela 7 – Melhores classificadores para classe OS	125
Tabela 8 – Resultados de trabalhos de detecção automática de crises. Adaptado de (ACHARYA <i>et al.</i> , 2013a).	126
Tabela 9 – Compilação dos resultados do gráfico de visibilidade para os trabalhos de Supriya <i>et al.</i> (2016), Wang <i>et al.</i> (2016) e Liu <i>et al.</i> (2016).	128
Tabela 10 – Seleção de atributos do grafo de visibilidade da classe NS	190

LISTA DE ABREVIATURAS E SIGLAS

ANN	Redes neurais artificiais
ANOVA	<i>One-way analysis of variance</i>
ApEn	<i>Approximate entropy</i>
AR	Auto-regressivos
ARI	Índice <i>Rand</i> ajustado
ARIMA	Auto-regressivos integrados e de médias móveis
ARMA	Auto-regressivos e de médias móveis
AUC	Área sobre a curva ROC
BA	Modelo Barabási-Albert ou rede livre de escala
CD	<i>Correlation dimension</i>
CSV	<i>Comma-separated value</i>
DT	Árvores de decisão
DWT	<i>Discrete Wavelet Transform</i>
EEG	Electroencefalograma
EMD	<i>Empirical mode decomposition</i>
ER	Modelo Erdős-Rényi ou rede aleatória
FFT	Transformada rápida de <i>Fourier</i>
FN	Falso negativo
FP	Falso positivo
FWHV	Grafo de visibilidade ponderado rápido
GANN	<i>Gaussian-based artificial neural network</i>
GMM	<i>Gaussian Mixture Model</i>
HOS	<i>Higher order spectra</i>
ICA	<i>Independent Component Analysis</i>
Knn	K vizinhos mais próximos
LDA	<i>Linear Discriminant Analysis</i>
LLE	<i>Largest Lyapunov exponent</i>
LS-SVM	<i>Least squares support vector machine</i>
MA	Médias móveis
NaN	<i>Not-a-Number</i>
NBC	<i>Naive Bayes Classifier</i>

PCA	<i>Principal Component Analysis</i>
PNN	<i>Probabilistic Neural Network</i>
RBF	<i>Radial basis function</i>
RFE	Eliminação de características recursiva
RFECV	Classificação de características pela eliminação recursiva de características e validação cruzada do número de melhores características
ROC	<i>Receiver operating characteristics</i>
RQA	<i>Recurrence Quantification Analysis</i>
SampEn	<i>Sample Entropy</i>
SNN	<i>Spiking neural network</i>
SVM	Máquina de vetores de suporte
TN	Verdadeiro negativo
TP	Verdadeiro positivo
WANN	<i>Wavelet-based artificial neural network</i>
WS	Modelo Watts-Strogatz ou rede pequeno mundo

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Contexto	21
1.2	Motivação	24
1.3	Hipótese e Objetivos	26
1.3.1	<i>Hipótese</i>	26
1.3.2	<i>Objetivos</i>	26
2	REVISÃO BIBLIOGRÁFICA	29
2.1	Epilepsia	29
2.1.1	<i>Eletroencefalograma</i>	29
2.1.2	<i>Caracterização</i>	30
2.1.3	<i>Classificação e detecção</i>	31
2.2	Séries temporais	35
2.2.1	<i>Caracterização</i>	35
2.2.2	<i>Análise de estacionariedade</i>	37
2.3	Conversão de séries temporais em redes complexas	39
2.3.1	<i>Redes complexas</i>	39
2.3.2	<i>Métricas para caracterização de redes complexas</i>	40
2.3.3	<i>Outros conceitos de redes complexas</i>	45
2.3.4	<i>Métodos de conversão de séries temporais em redes complexas</i>	46
2.3.5	<i>Análise de séries temporais por meio de redes complexas</i>	49
2.4	Considerações finais	52
3	MATERIAIS E MÉTODOS	53
3.1	Materiais	53
3.2	Métodos	56
3.2.1	<i>Estacionariedade</i>	56
3.2.1.1	<i>Estocástica</i>	58
3.2.1.2	<i>Determinística</i>	59
3.2.1.3	<i>Avaliação</i>	64
3.2.2	<i>Classificação de crises epiléticas</i>	65
3.2.2.1	<i>Leitura da série temporal</i>	66
3.2.2.2	<i>Conversão da série em grafo</i>	67

3.2.2.3	<i>Extração das métricas</i>	68
3.2.2.4	<i>Pré-processamento</i>	69
3.2.2.5	<i>Classificação</i>	70
3.2.2.6	<i>Avaliação</i>	73
4	RESULTADOS	75
4.1	Análise de estacionariedade	75
4.2	Classificação de crises epiléticas	79
4.2.1	<i>Conversão da série em grafo</i>	<i>80</i>
4.2.2	<i>Extração das métricas</i>	<i>81</i>
4.2.3	<i>Pré-processamento</i>	<i>82</i>
4.2.4	<i>Classificação</i>	<i>82</i>
4.2.5	<i>Avaliação</i>	<i>102</i>
4.2.6	<i>Comparação com a literatura</i>	<i>125</i>
5	CONCLUSÃO	129
5.1	Limitações	132
5.2	Trabalhos futuros	132
5.3	Contribuições	133
REFERÊNCIAS		135
APÊNDICE A	ANÁLISE DE VÉRTICES E ARESTAS	147
APÊNDICE B	EXTRAÇÃO DAS MÉTRICAS	181
APÊNDICE C	SELEÇÃO DE ATRIBUTOS	189

INTRODUÇÃO

1.1 Contexto

Epilepsia é uma desordem crônica do sistema nervoso central que predispõe o indivíduo a experimentar crises recorrentes. A espontaneidade e recorrência das crises é um aspecto comum de todas as epilepsias e são caracterizadas por descargas de alta-frequência sincronizadas de populações de neurônios do cérebro ([TEJADA *et al.*, 2013](#)). Epilepsia não é uma única doença, mas uma família de síndromes que compartilham a recorrência de crises. Os sintomas das crises vão desde um lapso na atenção a alucinações sensoriais, e até convulsão do corpo todo. Epilepsia pode desenvolver-se como resultado de um trauma no cérebro, um acidente vascular cerebral (AVC), uma infecção no cérebro ou até mesmo uma mutação do mecanismo molecular que regula o comportamento, migração ou organização dos neurônios ([SHOEB, 2009](#)).

Segundo um relatório na Organização Mundial da Saúde de 2009, atualmente, aproximadamente 50 milhões de pessoas têm epilepsia no mundo, e é estimado que 2,4 milhões são diagnosticados todo ano ([WHO, 2017](#)). Nos Estados Unidos da América, a epilepsia afeta 3 milhões de pessoas, sendo esta a terceira condição neurológica mais comum ([CHANG; LOWENSTEIN, 2003](#)). O objetivo do entendimento de como a epilepsia se desenvolve é sua prevenção.

O exame de Electroencefalografia (EEG), cuja descoberta é creditada ao psiquiatra alemão Hans Berger na década de 1920, mede os potenciais elétricos do cérebro. Seu potencial uso na epilepsia tornou-se claro assim que Gibbs e colegas em Boston identificaram um padrão de ondas característico de um tipo de epilepsia em 1936. Ainda hoje, o EEG continua com um papel central no diagnóstico e gerenciamento de pacientes com crises. Seu baixo custo e facilidade de aquisição, tornam evidentes as manifestações fisiológicas corticais anormais que estão por trás da epilepsia ([SMITH, 2005](#)).

O EEG é o processo de gravar a atividade elétrica cerebral através de vários eletrodos

(canais). Cada eletrodo captura a atividade elétrica de uma população de neurônios. Quando a atividade elétrica é medida através de eletrodos não invasivos, os eletrodos são colocados no couro cabeludo da pessoa. Outra maneira de se gravar a atividade elétrica é colocando os eletrodos na superfície do cérebro ou dentro do cérebro, nesse caso referido como EEG intracraniano ou eletrocorticograma (ECoG) para os eletrodos situados na superfície do cérebro.

Um possível mau posicionamento dos eletrodos invasivos, introduz erros de amostragem, portanto eles devem ser utilizados depois de uma exaustiva avaliação não invasiva. O EEG invasivo está associado a riscos adicionais apenas justificados se há uma boa chance de informação quanto à localização da zona epileptogênica. Raramente, descargas de forma epilética são gravadas em pessoas saudáveis, particularmente os jovens. O vídeo do momento de crise (vídeo ictal), e o EEG são considerados críticos na localização da zona epileptogênica. Uma análise cuidadosa dos primeiros sinais e sintomas da crise e de sua evolução fornecem importantes pistas sobre sua localização (NOACHTAR; RéMI, 2009).

A principal dificuldade à detecção de crises tanto para o EEG quanto para o ECoG é a variabilidade de pessoa para pessoa, tanto na crise quanto fora dela (SMITH, 2005; SHOEB, 2009). A localização e o conteúdo espectral do EEG variam de um indivíduo para outro. Outra característica do EEG que complica a detecção é a propriedade física do sinal: o EEG é mais sensível à atividade de populações de neurônios na superfície do cérebro. Consequentemente, a atividade de neurônios em estruturas dentro do cérebro tem pouca influência no EEG. Quando uma crise tem sua origem em uma estrutura dentro do cérebro o EEG pode refletir sequelas físicas da crise, como movimentos repetitivos de piscar o olho ou contração muscular, antes de refletir o hypersincronismo da atividade neural. Este tipo de crise é difícil de detectar com alta especificidade e baixa latência pois atividades como a do olho e contrações musculares são rotineiramente observadas como atividades diárias do indivíduo. O EEG está ainda sujeito a contaminações por fontes não fisiológicas. O balanço nos cabos dos eletrodos, alterações na interface do eletrodo com a pele, e interferências da rede elétrica podem produzir mudanças espectrais que afetam o desempenho do sinal do EEG (SHOEB, 2009).

As principais métricas de avaliação dos métodos de classificação de crises epiléticas utilizam a sensibilidade, a especificidade e a acurácia. Sensibilidade, também chamada de taxa de verdadeiros positivos, é a razão de instâncias que foram corretamente classificadas como da classe positiva. Especificidade, taxa de verdadeiros negativos, é a razão de instâncias que foram corretamente classificadas como da classe negativa. A acurácia é a taxa total de acertos pelo total de instâncias.

Alguns trabalhos fazem estudos comparativos de trabalhos anteriores (SMITH, 2005; MORMANN; ELGER; LEHNERTZ, 2006; GUO *et al.*, 2010; ACHARYA *et al.*, 2012a; ACHARYA *et al.*, 2013b). Os trabalhos mais recentes fazem essa comparação pela acurácia, e os trabalhos relatados em geral apresentam alta acurácia, entre 90% e 100%. Embora Smith (2005) mostre que em geral há alta especificidade, a sensibilidade é baixa no sinal de EEG para

detecção de crise.

Deve-se ter em mente que a acurácia deve ser analisada com cautela. Como em geral o conjunto de dados é desbalanceado, se os dados brutos foram analisados, há um período maior do sinal em um estado fora de crise. Esse desbalanceamento se não levado em conta pode mascarar o real desempenho do classificador, fazendo no pior caso que acerte quase sempre na classificação do estado fora da crise mas errando na classificação do estado de crise e ainda assim apresentando um bom valor de acurácia.

Nem toda crise apresenta sintomas tão visíveis quanto uma crise tônico-clônica, em algumas crises a pessoa pisca um pouco mais que o normal. Na crise de ausência, por exemplo, a pessoa simplesmente olha para o vazio e após alguns segundos volta a atividade que estava fazendo sem perceber que teve uma crise. Essa diversidade de sintomas das várias crises de epilepsia dificultam a tarefa de detecção. A possibilidade de detectar uma crise pelo EEG antes do aparecimento de sintomas clínicos é de grande importância, pois pode dar tempo para que uma ação seja tomada. Administração de um medicamento, ou um sinal de alerta para a pessoa, assim ela pode parar o carro caso esteja dirigindo ou fazendo alguma atividade que pode colocá-la em risco caso sofra uma crise.

A detecção de uma crise tipicamente é feita extraindo-se características do sinal, e as utilizando como entrada de um algoritmo de classificação. Alguns dos classificadores utilizados são redes neurais artificiais (ANN), máquina de vetores de suporte (SVM), árvores de decisão (DT) dentre outros. Algumas das características extraídas do sinal neural são coeficientes de transformada *wavelet*, entropias, coeficientes de modelos estatísticos como os auto-regressivos, dentre outras (ZHU; LI; WEN, 2014; GUO *et al.*, 2010; SHARMA; PACHORI, 2015; ACHARYA *et al.*, 2012a).

No contexto de processamento de sinais neurais, alguns trabalhos já utilizam o conceito de redes complexas para essa tarefa. Esta é uma abordagem recente e tem apresentado resultados interessantes (SCHINDLER *et al.*, 2008; ORTEGA; SOLA; PASTOR, 2008; RUBINOV; SPORNS, 2010; STRAATEN; STAM, 2013; AGGARWAL; SUBBIAN, 2014; LEHNERTZ *et al.*, 2014).

A grande disponibilidade de dados possibilitou o mapeamento e estudo da topologia de várias redes, já que redes complexas surgem em vários sistemas artificiais e naturais. O cérebro consiste da interconexão de neurônios, inclusive uma característica topológica de redes corticais é a chamada rede livre de escala, organização recorrente em várias outras redes complexas já estudadas (FREEMAN; BREAKSPEAR, 2007). Na busca pelas leis que governam a dinâmica e evolução desses sistemas deu-se início a uma análise sistemática e caracterização dessas redes.

Uma rede, em sua forma mais simples, é uma coleção de pontos unidos por pares de linhas. Muitos objetos de interesse na física, biologia, e ciências sociais podem ser pensados como redes. Muitos sistemas de interesse aos cientistas são compostos por partes individuais e

componentes ligados de alguma maneira. Muitas pessoas estudam a natureza dos componentes individuais, como um computador funciona ou como o ser humano sente e age, enquanto outros estudam a natureza das conexões ou interações, os protocolos de comunicação utilizados na Internet ou a dinâmica das amizades. Mas há um terceiro aspecto desses sistemas interativos, às vezes negligenciados mas quase sempre cruciais ao comportamento do sistema, que são os padrões de conexões entre os componentes (NEWMAN, 2010). O padrão das conexões em um dado sistema pode ser representado como uma rede, os componentes do sistema sendo a rede de vértices e as conexões, as arestas. Com um pouco de reflexão não deve ser surpresa que a estrutura de tais redes, o padrão particular de interações, pode ter um grande efeito no comportamento do sistema (NEWMAN, 2010).

Neste contexto a abordagem da física estatística foi explorada como uma estratégia muito conveniente por causa de sua forte conexão com a teoria de estatística em grafos e a possibilidade de caracterizar a emergência de fenômenos macroscópicos em termos da evolução dinâmica dos elementos básicos do sistema. Além disso, a análise em larga escala de redes aparecendo em vários campos diferentes proveu evidência para o compartilhamento de várias propriedades assintóticas e levantou a questão da suposta emergência nas redes complexas de princípios de organização comum e geral que vai além dos sistemas individuais e particulares, e são assim passíveis de princípios de modelagem geral (BARRAT; BARTHELEMY; VESPIGNANI, 2008).

Uma vez que a inspeção visual de um gráfico de uma rede, seu grafo, não permite extrair muitas informações úteis, especialmente para redes grandes, várias medidas foram desenvolvidas ou adaptadas para a caracterização de redes. Métricas como menores caminhos, coeficiente de aglomeração, distribuição do grau, medidas de correlação, dentre outras, são utilizadas para a descrição da ordenação estrutural e hierárquica das redes.

Redes são um meio geral mas ainda assim poderoso, de representar padrões de conexões ou interações entre as partes do sistema. Este é um grande motivador do estudo de sistemas através de sua representação em rede. Em suma, o estudo de redes complexas é uma ferramenta para a análise de sistemas dinâmicos, como a dinâmica dos sinais neurais durante uma crise epilética (ZHU, 2012; SUBRAMANIAM; HYTTINEN, 2013; LEHNERTZ *et al.*, 2014; ZHU; LI; WEN, 2014).

Um passo natural seria, então, a utilização de redes complexas para análise e modelagem de séries temporais de algum sistema dinâmico como o cérebro por exemplo.

1.2 Motivação

Na literatura pesquisada percebe-se o uso crescente e com bons resultados de redes complexas para a análise de séries temporais (ZHANG; SMALL, 2006; LI; WANG, 2007; REIJNEVELD *et al.*, 2007; ZHANG *et al.*, 2008; XU; ZHANG; SMALL, 2008; SCHINDLER *et al.*, 2008; LACASA *et al.*, 2008; SHIMADA; KIMURA; IKEGUCHI, 2008; YANG; YANG,

2008; ORTEGA; SOLA; PASTOR, 2008; FERRI *et al.*, 2008; BULLMORE; SPORNS, 2009; ELSNER; JAGGER; FOGARTY, 2009; MARWAN *et al.*, 2009; SHIRAZI *et al.*, 2009; HORSTMANN *et al.*, 2010; PONTEN *et al.*, 2010; RUBINOV; SPORNS, 2010; AMINI *et al.*, 2010; SHAO, 2010; DONNER *et al.*, 2011; YANG; LIU; LIU, 2012; NUÑEZ *et al.*, 2012a; NUÑEZ *et al.*, 2012b; DONGES; DONNER; KURTHS, 2013; STRAATEN; STAM, 2013; AHADPOUR; SADRA; ARASTEHFARD, 2014; AGGARWAL; SUBBIAN, 2014; ZHU; LI; WEN, 2014; LEHNERTZ *et al.*, 2014), mas encontrou-se poucos trabalhos que utilizam redes complexas para analisar a dinâmica da crise epilética (REIJNEVELD *et al.*, 2007; SCHINDLER *et al.*, 2008; ZHU, 2012; SUBRAMANIAM; HYTTINEN, 2013; LANG *et al.*, 2013). Normalmente analisa-se o sinal em termos da topologia e conectividade dos eletrodos e medidas de correlações entre os sinais dos mesmos, mas não a dinâmica da crise caracterizada através de redes complexas (SCHINDLER *et al.*, 2008). Há aqui então, uma lacuna no estudo de epilepsia utilizando uma ferramenta promissora para se analisar séries temporais.

Alguns trabalhos citam a incerteza da predição de crises utilizando algoritmos prospectivos (MORMANN; ELGER; LEHNERTZ, 2006; LEHNERTZ *et al.*, 2007; LEHNERTZ *et al.*, 2009; CARNEY; MYERS; GEYER, 2011). Incerteza devido à falta de rigor na análise dos resultados e aplicação dos métodos. Fato que contribui para a constatação de vários trabalhos com resultados contraditórios devido à negligência com o rigor estatístico (LEHNERTZ *et al.*, 2007).

O desenvolvimento de técnicas de predição eficientes, juntamente com estratégias de tratamento de drogas liberadas antes de uma crise, podem reduzir os efeitos colaterais como sedação e pensamento anuviado. Tais sistemas podem ainda diminuir consideravelmente a morbidade e mortalidade assim como aumentar a qualidade de vida das pessoas com epilepsia (CARNEY; MYERS; GEYER, 2011). Se alguma técnica de predição mostrar-se particularmente útil em vários cenários, é possível o desenvolvimento de um equipamento que monitora o EEG de uma pessoa a fim de tomar uma ação quando uma crise for detectada. Devido ao baixo custo do EEG, se um equipamento de baixo custo com bom desempenho for desenvolvido, ONGs ou o governo poderiam realizar mutirões para realizar o diagnóstico precoce na população em geral.

Outra constatação da literatura sobre esse tema é a falta de trabalhos completos sobre os métodos de conversão de séries temporais em redes complexas e sua caracterização. O presente trabalho pretende, portanto, utilizar vários desses métodos de conversão e experimentar diversos valores para seus parâmetros a fim de cobrir bem o espaço das possíveis conversões, e assim estudar quais técnicas de conversão, e parâmetros, apresentam bom desempenho na classificação.

Como muitos trabalhos analisam apenas a diferença entre um sinal de pessoa saudável e um sinal durante a crise, e tal cenário não oferece uso prático, este trabalho pretende investigar vários cenários de interesse prático. Alguns desses cenários são a distinção entre um sinal de pacientes com epilepsia dentro e fora da crise, e a distinção de um sinal de paciente com epilepsia fora da crise e o sinal de uma pessoa saudável. Esses cenários são de interesse prático, pois

pensando o desenvolvimento de um dispositivo que consiga fazer uma boa distinção em tais cenários torna-se adequado para uso clínico, diferentemente dos cenários de muitos trabalhos apresentados na literatura.

Ainda há muito a se descobrir sobre a epilepsia, e uma nova abordagem, como redes complexas, pode prover novos *insights* quanto à epileptogênese.

1.3 Hipótese e Objetivos

Esta seção apresenta a hipótese e os objetivos deste trabalho.

1.3.1 Hipótese

O uso de redes complexas possibilita o desenvolvimento de um algoritmo prospectivo de predição de crises epiléticas com desempenho comparável ao estado da arte, e estatisticamente melhor, mais acurado e com melhor sensibilidade e especificade, que um preditor aleatório.

1.3.2 Objetivos

A fim de confirmar a hipótese desse trabalho pretende-se atingir os seguintes objetivos.

- **Objetivo geral:**

- Estudo completo da predição de crises epiléticas utilizando métodos de conversão de séries temporais em redes complexas.

- **Objetivos específicos:**

- Estudo extensivo dos algoritmos de conversão de séries temporais em redes complexas;
- Comparação do desempenho dos algoritmos de predição utilizando redes complexas com os métodos do estado da arte na predição de epilepsia, utilizando um *framework* estatístico sólido e rigoroso;
- Avaliação do desempenho em vários conjuntos de dados independentes, das bases de dados mais utilizadas na literatura bem como de gravações de longo período.
- Estudo das abordagens de predição utilizando sinais de vários cenários, e não apenas entre sinal de pessoas saudáveis e de pessoas com epilepsia durante uma crise.

A figura 1 traz o esquema geral do trabalho para cumprir os objetivos. Primeiramente é feita a leitura da série temporal, então cada janela da série é convertida em grafo pelos vários métodos de conversão. Com o grafo construído, as métricas descritas são extraídas, e um conjunto de dados na forma atributo-valor é criado. Este conjunto é pré-processado, e então os algoritmos

de classificação são utilizados gerando as matrizes de confusão. Calcula-se então as métricas de desempenho e a avaliação final é realizada.

Figura 1 – Etapas utilizadas na classificação.

Pretende-se estudar o sinal em 5 cenários diferentes: sinal normal, de pessoa saudável, e durante uma crise; sinal interictal, sinal de pessoa com epilepsia mas fora da crise, de eletrodo fora do foco da crise e sinal durante uma crise; sinal interictal de eletrodo no foco da crise e sinal durante a crise; sinal interictal das regiões focal e não-focal, e sinal normal; sinal interictal das regiões focal e não-focal, e sinal durante uma crise.

Em cada cenário serão realizadas 75 conversões diferentes, variando os parâmetros de cada método de conversão. Em cada conversão 7 valores para tamanhos de janelas serão utilizados. Duas medidas de desempenho serão avaliadas. E em cada conversão, 37 classificadores diferentes serão induzidos a fim utilizar uma variedade de classificadores.

REVISÃO BIBLIOGRÁFICA

2.1 Epilepsia

Esta seção traz uma descrição sobre epilepsia, e alguns aspectos a ela relacionados necessários ao entendimento do trabalho.

2.1.1 Eletroencefalograma

O EEG, como dito anteriormente, é uma importante ferramenta para o diagnóstico e pesquisa de epilepsia. O neurologista *William Gordon Lennox* foi pioneiro no diagnóstico de epilepsia usando o EEG ([LENNOX, 1946](#) apud [MEDEIROS; MORAES, 2014](#)).

O psiquiatra alemão Hans Berger fez a primeira gravação de um campo elétrico do cérebro humano em 1924 na Alemanha. Ele deu o nome a esta gravação de eletroencefalograma (EEG). A atividade espontânea é medida no couro cabeludo ou na superfície do cérebro. A amplitude do EEG é de cerca de 100 μV quando medida no couro cabeludo, e cerca de 1 à 2 mV quando medida na superfície do cérebro ([BERGER, 1929](#) apud [MALMIVUO; PLONSEY, 1995](#)).

Geralmente o sistema internacionalmente padronizado 10-20 de posicionamento é empregado para gravações de EEG espontâneo. Neste sistema, os eletrodos são colocados na superfície do couro cabeludo, e suas localizações são determinadas dividindo-se o perímetro em intervalos de 10% e 20%. A localização e nomenclatura desses eletrodos é padronizada pela Sociedade Americana de Eletroencefalografia. Além do sistema 10-20, existem outros esquemas de localização de eletrodos para gravação de atividade neural ([MALMIVUO; PLONSEY, 1995](#)). Como as bases disponíveis em geral utilizam esse sistema, inclusive as empregadas neste trabalho, os outros sistemas de posicionamento não serão descritos.

A figura 2 mostra uma configuração de eletrodos no sistema 10-20 e sua respectiva

nomenclatura.

Figura 2 – Sistema 10-20 de posicionamento de eletrodos ([WIKIPEDIA, 2016](#)).

Os atuais equipamentos de aquisição de EEG convertem o sinal analógico do eletrodo para digital e também aplicam filtros a fim de remover fontes de interferência como a rede elétrica. E após esses passos, o sinal, agora digital, é armazenado em um computador ou outro dispositivo eletrônico permitindo a análise posterior mais detalhada desses dados.

2.1.2 Caracterização

[Chang e Lowenstein \(2003\)](#) fazem uma revisão sobre epilepsia destacando as ideias dominantes sobre a epileptogênese. Esse trabalho, que é voltado para a área médica, descreve os mecanismos moleculares e celulares da epilepsia, questões que não serão abordadas no presente trabalho. Esse trabalho traz também, algumas classificações e considerações que podem mostrar-se importantes para o desenvolvimento de conclusões. De acordo com esse trabalho, epilepsia engloba várias síndromes cuja principal característica é a predisposição a crises (do inglês *seizures*) recorrentes não provocadas.

Uma dada crise pode ser classificada por sua característica clínica, por exemplo, uma crise complexa parcial ou uma crise generalizada tônico-clônica. As síndromes são divididas em basicamente duas categorias: síndromes generalizadas ou parciais (localizadas).

Na síndrome generalizada, o tipo predominante de crise tem origem nos dois hemisférios cerebrais simultaneamente; tipicamente há um forte fator genético neste caso. Na síndrome parcial, as crises se originam em um ou mais focos e podem espalhar-se para o cérebro inteiro; a maior parte deste tipo de síndrome acredita-se que seja resultado de um ou mais traumas no sistema nervoso central.

A crise de ausência (do inglês *absence seizure*), um tipo de crise generalizada, faz com que a pessoa olhe para o vazio e cesse sua atividade normal por alguns segundos e então retome

suas atividades sem memória do ocorrido. Há um padrão clássico no EEG que caracteriza essa crise, chamado de *three-per-second spike-wave discharge*. Sabe-se muito pouco sobre o processo por trás da desordem parcial, apesar dessa ser o tipo mais comum de crise em adultos. Normalmente tem origem em lesões localizadas como traumas na cabeça, acidente vascular cerebral (AVC) e tumores. A origem mais comum deste tipo de síndrome é no córtex temporal mesial.

Período ictal é o nome dado ao estado fisiológico quando uma crise ocorre, e interictal é o período entre crises epiléticas. [Staley e Dudek \(2006\)](#) afirmam que os picos interictais (do inglês *interictal spikes*) são amplamente aceitos como sinal de epilepsia. Estes são picos breves, duram menos de 250 milissegundos e são morfológicamente definidos, observados no EEG de pacientes com epilepsia localizada. Esses picos se originam da sincronização de descargas de um grupo de neurônios localizados no foco epilético.

Embora o EEG seja a forma mais comum de diagnóstico de epilepsia, em alguns pacientes, apenas o EEG no couro cabeludo não é suficiente para localizar muito bem o foco epilético. Nesses casos é preciso um procedimento invasivo de obtenção da atividade neural, posicionando eletrodos na superfície do cérebro ou implantando-os dentro do cérebro. Mas como esses procedimentos incorrem em maiores riscos para o paciente, devem ser utilizados após extensivo uso do EEG no couro cabeludo.

Devido a proximidade do eletrodo à população neural que está sendo medida, o EEG invasivo apresenta algumas vantagens: sua relação sinal-ruído é melhor, o registro de artefatos¹ é consideravelmente reduzido, e a sensibilidade é ampliada, enquanto que o EEG no couro cabeludo pode não registrar uma crise leve, o invasivo pode conseguir registrá-la.

[Kramer e Cash \(2012\)](#) dividem a evolução de uma crise em 3 estágios: começo, propagação e terminação. Em cada estágio descrevem o ritmo neural e a sincronização específicos, e tais características podem ser usadas na detecção de crises. Embora essa distinção seja interessante e possa até mesmo melhorar a tarefa de identificação de crise, como os conjuntos de dados disponibilizados, tipicamente não rotulam os dados distinguindo essas etapas da crise; na prática a única informação utilizada é segmento do sinal em que a crise ocorre.

2.1.3 Classificação e detecção

É apresentada nessa seção como em parte da literatura é feita a classificação e detecção de sinais epiléticos, os métodos e métricas utilizados.

O método geral dos trabalhos segue a figura 3. Primeiramente o sinal bruto, sem pré-processamento, do EEG é carregado, depois ele passa por um processo de extração de caracte-

¹ Artefato é o nome dado à atividade elétrica que contamina o EEG. Um exemplo de artefato é o sinal elétrico devido a ativação de um músculo. Como o EEG no couro cabeludo é gravado com eletrodos posicionados na pele sua proximidade com o músculo é maior do que com as populações neurais sendo medidas ([NOACHTAR; RÉMI, 2009](#)).

terísticas, dependendo do sinal pode ser feita a extração de artefatos presentes. Em seguida, essas características extraídas são inseridas em um algoritmo de classificação e após a saída do classificador são feitos os cálculos estatísticos para a validade da solução encontrada.

Figura 3 – Método para classificação de EEG. Adaptado de (ZHU; LI; WEN, 2014).

Na bibliografia pesquisada a remoção dos artefatos foi feita por inspeção visual (KAN-NATHAL *et al.*, 2005a; GUO *et al.*, 2010), *Independent Component Analysis* (ICA) (ACHARYA *et al.*, 2013b; GUERRERO-MOSQUERA *et al.*, 2010), filtragem (MIROWSKI *et al.*, 2008; GUERRERO-MOSQUERA *et al.*, 2010) ou anulação dos valores da janela do sinal sendo analisado se o desvio-padrão dessa janela ultrapassar um dado limiar (BERGSTROM *et al.*, 2013). Apesar dos trabalhos citados explicitarem como a remoção de artefatos foi feita, a maioria não o faz. Assume-se que os dados utilizados estão livres de artefatos, ou então seu extrator de características ou seu método de classificação incorporam os artefatos no processamento.

Mormann, Elger e Lehnertz (2006) afirmam que a maioria das técnicas de predição tem certas características em comum, como uso de janelas móveis na qual uma medida linear ou não linear é calculada para cada janela e em seguida move-se para a próxima janela. O tamanho da janela varia de 10 a 40 segundos e as análises usam um canal do sinal (análise univariada), ou relações entre vários (análise multivariada).

Também afirmam que na avaliação dos resultados, se uma crise não for precedida por um alarme dentro de um horizonte de previsão, um falso positivo é computado. A sensibilidade do algoritmo de predição é então quantificada como o número de crises com pelo menos um alarme dentro do horizonte de previsão, dividido pelo total de crises presentes.

Para avaliar estatisticamente a especificidade do algoritmo de predição, muitos estudos reportam a taxa de especificidade medida em falsas predições por hora. É preciso atenção pois os trabalhos não utilizam a mesma definição de falso positivo. Alguns trabalhos contam todos os falsos positivos e dividem pela duração total da gravação.

Uma maneira melhor de avaliar a especificidade da predição seria reportar a porção do tempo do período interictal, entre crises, durante o qual o paciente não está em um estado de

espera de uma falsa crise (falso positivo), ou seja, a crise não acontecerá.

Mormann, Elger e Lehnertz (2006) concluem que a literatura por eles descrita não permite uma definição conclusiva se crises são predizíveis usando algoritmos prospectivos. E que para uma resposta definitiva a essa questão uma metodologia mais rígida incluindo uma validação estatística rigorosa é necessária. Algoritmos prospectivos são aqueles que para a predição de séries temporais, utilizam informações presentes até o momento, ou seja, não utilizam amostras ou outras informações futuras para a predição, como saber em qual canal a crise está localizada.

Outra revisão, feita por Lehnertz *et al.* (2007), mostra que muitos estudos são negligentes na avaliação de desempenho de suas técnicas, o que colabora para descobertas contraditórias. Segundo os autores, há duas abordagens para avaliar o desempenho, uma abordagem analítica, e outra baseada em *bootstrap*. Ambas vão além do simples cálculo da especificidade e sensibilidade. A abordagem analítica faz uso da chamada característica de predição de crise (do inglês *seizure prediction characteristic*) e para uma dada especificidade compara a sensibilidade do método sendo investigado com o desempenho de um preditor não especificado ingênuo, aleatório ou periódico (do inglês *naive predictor*).

Lehnertz *et al.* (2007), assim como Mormann, Elger e Lehnertz (2006), afirmam que o estado da arte ainda não permite uma resposta definitiva à questão se crises podem ser preditas usando algoritmos prospectivos.

A maioria dos trabalhos na literatura de classificação de sinais de epilepsia utiliza o conjunto de dados fornecido pela Universidade de Bonn, da Alemanha (ANDRZEJAK *et al.*, 2001), os demais utilizam dados próprios ou não mais disponíveis. Essa base de dados será descrita na seção que trata dos materiais utilizados neste trabalho.

A lista abaixo apresenta as técnicas de extração de características do sinal utilizadas juntamente com a bibliografia pesquisada onde a mesma foi utilizada:

- Transformada *Wavelet*, harmônica, *Daubechies* e multi-resolução (ADELI; ZHOU; DADMEHR, 2003; ADELI; GHOSH-DASTIDAR; DADMEHR, 2007; OCAK, 2009; GUO *et al.*, 2010);
- Entropias, *Shannon's entropy*, *Renyi's entropy*, *Kolmogorov—Sinai entropy*, *approximate entropy* (ApEn), *Sample Entropy* (SampEn), *Phase Entropy 1 - S1*, and *Phase Entropy 2 - S2* (KANNATHAL *et al.*, 2005a; OCAK, 2009; ACHARYA *et al.*, 2012a);
- Medidas de sistemas não-lineares, *correlation dimension* (CD) e *largest Lyapunov exponent* (LLE) (ADELI; GHOSH-DASTIDAR; DADMEHR, 2007);
- *Line length* (GUO *et al.*, 2010);
- *Empirical mode decomposition* (EMD) (SHARMA; PACHORI, 2015).

Abaixo são listados os classificadores utilizados para classificação de crises bem como a referência onde o referido classificador foi empregado:

- SVM, *Least squares support vector machine* (LS-SVM), com os seguintes *kernels*, *radial basis function* (RBF), *Mexican hat wavelet kernel* e *Morlet wavelet kernel* (MIROWSKI *et al.*, 2008; ACHARYA *et al.*, 2012a; SHARMA; PACHORI, 2015);
- Redes neurais artificiais, *Gaussian-based artificial neural network* (GANN), *wavelet-based artificial neural network* (WANN), *Spiking neural network* (SNN), *convolutional networks*, *Probabilistic Neural Network* (PNN) (CHIU *et al.*, 2005; GHOSH-DASTIDAR; ADELI, 2007; MIROWSKI *et al.*, 2008; GUO *et al.*, 2010; ACHARYA *et al.*, 2012a);
- Regressão, *L1-regularized logistic regression* (MIROWSKI *et al.*, 2008);
- Árvores de decisão (ACHARYA *et al.*, 2012a);
- *Gaussian Mixture Model* (GMM) (ACHARYA *et al.*, 2012a);
- *Naive Bayes Classifier* (NBC) (ACHARYA *et al.*, 2012a).

Por fim são listadas as métricas e técnicas de avaliação utilizadas e os trabalhos que as utilizaram:

- Sensitividade (MORMANN; ELGER; LEHNERTZ, 2006; MIROWSKI *et al.*, 2008; GUO *et al.*, 2010; SHARMA; PACHORI, 2015);
- Especificidade (GUO *et al.*, 2010; SHARMA; PACHORI, 2015);
- Acurácia (KANNATHAL *et al.*, 2005a; CHIU *et al.*, 2005; GHOSH-DASTIDAR; ADELI, 2007; OCAK, 2009; GUO *et al.*, 2010; ACHARYA *et al.*, 2012a; SHARMA; PACHORI, 2015);
- Teste estatístico t (KANNATHAL *et al.*, 2005a);
- *One-way analysis of variance* (ANOVA) (ADELI; GHOSH-DASTIDAR; DADMEHR, 2007);
- *Surrogate data analysis* (OCAK, 2009).

Um trabalho mais recente (CARNEY; MYERS; GEYER, 2011), concorda com Mormann, Elger e Lehnertz (2006), Lehnertz *et al.* (2007) de que os estudos de predição de crises não utilizam algoritmos prospectivos, apenas retrospectivos, ou seja, nenhuma evidência de utilidade clínica (algoritmo prospectivo, cego e aleatório) existe. Essa importante limitação é uma das principais dificuldades na ciência de predição de crises, e em parte é devida à falta de gravações de EEG uniformes de longo prazo e rigorosos testes estatísticos para especificidade e sensibilidade dos métodos.

2.2 Séries temporais

O EEG pode ser estudado como uma série temporal uma vez que é composto por amostras da tensão elétrica do cérebro ordenadas temporalmente. Portanto, o estudo de métodos de análise de séries temporais faz-se necessário.

2.2.1 Caracterização

Embora nos sinais de EEG tenha-se a suspeita de que haja um atrator caótico subjacente governando a dinâmica, esses sinais também podem ser modelado por funções aleatórias ou processos estocásticos.

Matematicamente, uma série temporal é definida como um conjunto X_t de amostras ordenadas, sendo t o índice da amostra. Uma abordagem tradicional no estudo de séries temporais é modelar cada amostra X_t como uma variável aleatória, ou seja, a família $\{X_t, t \in T\}$, onde T é o conjunto dos índices da série temporal, é um processo estocástico.

Morettin e Tolo (2006) relatam haver 4 objetivos para o estudo de séries temporais: investigar o mecanismo gerador da série, fazer previsões de valores futuros, descrever o comportamento da série e procurar periodicidades relevantes. O interesse deste trabalho, e em geral no estudo do EEG como uma série temporal, está em investigar o mecanismo gerador da série e em alguns casos descrever o comportamento da série. Sempre que se usa modelos para o estudo de fenômenos físicos, suposições são feitas a fim de permitir o tratamento analítico do modelo em questão. A classe de modelos auto-regressivos (AR), de médias móveis (MA), auto-regressivos e de médias móveis (ARMA) e auto-regressivos integrados e de médias móveis (ARIMA), recebeu grande atenção no estudo de séries temporais. O problema com esses modelos é que, com exceção do modelo ARIMA, eles supõem a estacionariedade da série. Esta é uma suposição forte e que pouco ocorre na prática, e mesmo o modelo ARIMA, que modela bem a série na presença de tipos de não estacionariedade, não é adequado em todos os cenários. Estacionariedade é um conceito importante, e será melhor discutida na seção 2.2.2.

O modelo AR é assim chamado pois os valores futuros são obtidos por meio de uma regressão de alguns valores passados, ou seja, uma auto-regressão. A expressão 2.1 mostra sua formulação matemática. Nela, um modelo AR de ordem p é especificado, sendo os coeficientes ϕ_i os pesos da auto-regressão, e a_t um termo de erro aleatório.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t \quad (2.1)$$

Por outro lado, o modelo MA representa a série temporal como uma combinação linear finita dos termos aleatórios a_t , sua especificação matemática é descrita na equação 2.2, que especifica um modelo MA de ordem q , sendo os coeficientes θ_i os pesos da combinação linear

dos a_j . O nome média móvel pode ser enganoso pois a soma dos termos $1, -\theta_1, -\theta_2, \dots, -\theta_q$ não necessariamente totaliza 1, mas essa é nomenclatura tipicamente utilizada para esse modelo.

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2.2)$$

Existe uma relação de equivalência entre os modelos AR e MA, uma série descrita por um modelo AR de ordem finita pode também ser especificada por um modelo MA de ordem infinita (infinitos parâmetros θ). O inverso também ocorre, uma série descrita por um modelo MA de ordem finita pode ser especificado por um modelo AR de ordem infinita (infinitos parâmetros ϕ).

É possível ainda combinar os modelos AR e MA no chamado modelo ARMA. Isso é útil em cenários em que a série não é completamente especificada por apenas um dos dois modelos garantindo maior flexibilidade do modelo ARMA. A equação 2.3 descreve um modelo ARMA de ordens p e q .

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2.3)$$

Os três modelos aqui descritos não são adequados se a série sendo modelada apresentar as chamadas não estacionariedades de nível e de inclinação. O primeiro caso acontece quando a série tem seu nível deslocado, por exemplo se ela estava centrada no zero mas por algum motivo passou a estar centrada no nível 2 (valor do eixo da ordenada igual a 2), isso causa problemas nos modelos anteriores. O segundo caso ocorre, quando as oscilações da série sofrem uma inclinação, ao invés das oscilações estarem na horizontal passam a seguir uma reta inclinada.

Nesses dois cenários o modelo ARIMA consegue descrever bem o comportamento da série. Ele incorpora ao modelo ARMA uma transformação à série original removendo a não estacionariedade permitindo que o modelo ARMA represente bem essa nova série. A equação do modelo ARIMA de ordem (p, d, q) é apresentada na equação 2.4. Ela representa um modelo ARMA da série Y_t que é obtida aplicando o operador de diferença finita ∇^d na série X_t .

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2.4)$$

A equação 2.5 mostra o operador ∇^1 , operador de diferença de ordem 1.

$$Y_t = \nabla^1 * X_t = X_t - X_{t-1} \quad (2.5)$$

A operação de diferença é a responsável pela remoção da mudança de nível e inclinação na série original X_t .

Os modelos acima descritos são paramétricos pois um número finito de parâmetros especifica completamente o modelo. Há também modelos chamados não-paramétricos, e estes precisam de um número infinito de parâmetros para ajustarem-se completamente à série sendo modelada. Exemplos de modelos não-paramétricos são a função de autocovariância da série e sua transformada de Fourier.

Um detalhamento mais profundo de modelos e métodos de análise de séries temporais é descrito por [Morettin e Toloi \(2006\)](#) e [Box, Jenkins e Reinsel \(2013\)](#).

Analizando-se os parâmetros desses modelos é possível descrever o comportamento da série e investigar seus mecanismos geradores, os quais eram os objetivos da análise de séries temporais citados inicialmente como tipicamente utilizados no estudo de sinais de EEG.

2.2.2 *Análise de estacionariedade*

No processo de revisão bibliográfica vários autores colocam a propriedade de estacionariedade como importante, sendo a mesma um requisito à aplicação de determinados métodos de análise. Mas raramente é descrito como tal propriedade foi checada nos dados sendo analisados, frequentemente, fazendo-se a suposição de que os dados são estacionários ou sem nem mesmo mencionar tal suposição. Visto isso, o autor começou uma investigação sobre as maneiras de aferir a estacionariedade. Esta seção traz a revisão sobre a análise de estacionariedade.

O estudo deste conceito aconteceu de uma investigação paralela ao objetivo principal desta dissertação. Ela se deu pois, como mencionado anteriormente, alguns métodos de análise pressupõem a estacionariedade mas ela não é checada, então começou-se uma investigação sobre esse conceito. Devido à importância do conceito e os resultados obtidos sobre ele serem interessantes, decidiu-se dedicar essa seção à esse tema.

Em geral, quando fala-se em análise de estacionariedade de um sinal ou sistema, costuma-se utilizar a definição estatística de estacionariedade. Esta definição especifica que um sinal é tido como estacionário se todos os momentos estatísticos, média (primeiro momento), variância (segundo momento), terceiro momento, etc, mantêm-se constantes ao longo do tempo. Esta é definida como estacionariedade forte, e tem pouco uso prático. Mas mesmo a chamada estacionariedade fraca, que relaxa as condições para um sinal ser definido como estacionário, na concepção do autor não é suficiente para capturar um conceito mais geral de estacionariedade. Isso é evidente quando fórmulas determinísticas são usadas para modelar sistemas dinâmicos.

O EEG, como já foi dito, é um método não invasivo e de baixo custo para a aquisição de um sinal que mede a atividade elétrica do cérebro. Uma das dificuldades de se analisar este tipo de sinal é a diversidade da dinâmica do cérebro a qual ainda não é bem entendida. O EEG pode ser estudado como uma série temporal, e séries temporais podem ser analisadas de duas perspectivas, estocástica e determinística ([ZOLDI; KRYSTAL; GREENSIDE, 2000](#); [BLANCO et al., 1995](#); [STAM, 2005](#)).

Uma classe de modelos estocástica para análise de séries temporais que teve bastante atenção, foram os chamados modelos estacionários. Eles assumem o equilíbrio estatístico, ou seja, as propriedades probabilísticas não mudam com o tempo; filtros lineares, modelos AR, MA e ARMA são alguns exemplos.

No cenário não estacionário, esses modelos perdem um pouco de sua interpretabilidade, pois, os parâmetros não representam mais o que foram pensados para expressar. Um exemplo são os coeficientes da transformada de *Fourier*, na presença de não estacionariedade eles se tornam difíceis de interpretar, já que é suposto que a energia em dada frequência está presente no sinal de maneira uniforme, o que não é o caso em sinais não estacionários (BOX; JENKINS; REINSEL, 2013; BLANCO *et al.*, 1995; BENDAT; PIERSOL, 2011).

O teorema de *Takens*, a fundamentação teórica da reconstrução do espaço de fase, requer que o sinal sendo analisado esteja numa trajetória assintótica ao redor do atrator do sistema (TAKENS, 1981). E isso dificilmente acontecerá durante transições devido à não estacionariedade. Já que os parâmetros do sistema mudam, possivelmente mudando o próprio atrator, leva tempo até que uma trajetória assintótica seja alcançada. O espaço de fase é uma representação de um sistema dinâmico na qual cada ponto representa um possível estado do sistema. Vários métodos de séries temporais usam este conceito no estudo de tais sistemas (STROGATZ, 2014).

A palavra *estacionário* significa permanecer parado, mantendo a posição. O entendimento mais genérico do conceito de estacionariedade é o de que os parâmetros do sistema não se alteram. Da perspectiva de sistemas dinâmicos sendo modelados pelas abordagens estocástica e determinística, a caracterização da mudança dos parâmetros também pode ser apresentada com essas duas perspectivas.

A visão estocástica, modela a série temporal como um processo estocástico, ou seja, cada elemento da série é uma variável aleatória amostrada de uma função de densidade de probabilidade. Esta visão tenta mensurar se a distribuição da variável aleatória muda com o tempo. Na análise de séries temporais, há duas definições de estacionariedade, a estacionariedade forte e a fraca. Estas são as definições típicas de estacionariedade, e na prática apenas a estacionariedade fraca é usada pois a forte implica que todos os momentos estatísticos da série sejam invariantes à translação temporal. Enquanto que na fraca, apenas o primeiro momento permanece constante e a função de autocorrelação é dependente apenas do deslocamento temporal (BOX; JENKINS; REINSEL, 2013; BENDAT; PIERSOL, 2011).

Do ponto de vista determinístico, a invariância dos parâmetros pode ser entendida como se a evolução do sistema mantivesse sua trajetória com o passar do tempo. É como se as equações que governam a trajetória do sistema permanecessem constantes. Uma maneira prática de determinar a estacionariedade é verificar se o sistema é autônomo ou não, ou seja, se o tempo tem alguma influência na equação de evolução do sistema (MANUCA; SAVIT, 1996).

Há vários aspectos que influenciam a percepção de mudança. Uma técnica comum

do estudo de séries temporais é o janelamento, ou seja, analisar a série através de uma janela deslizante. E o tamanho da janela é um componente chave na sensação de variação dos parâmetros. A medida usada para avaliar a evolução dos parâmetros, também tem grande importância na percepção de variação (MANUCA; SAVIT, 1996).

Enquanto o fenômeno sendo estudado pode gerar dados não estacionários, há maneiras de torná-los estacionários. Usando o operador de diferenças finitas, por exemplo, alguns tipos de séries não estacionárias podem tornar-se estacionárias, mas este método não é adequado em toda a situação. Considerando que toda amostra de série temporal tem um componente de erro seguindo uma distribuição *gaussiana*, então como a série transformada foi obtida subtraindo suas amostras, o componente de erro tem agora uma variância maior que da série original. Isto ocorre, pois na adição ou subtração de variáveis aleatórias, a variância sofre um aumento. Este aumento tem consequências danosas em vários métodos de análise, como na reconstrução do espaço de fase. Além disso, não há garantias de que a transformação manterá a estrutura subjacente do processo gerador da série intacta (BOX; JENKINS; REINSEL, 2013; THEILER; EUBANK, 1993).

2.3 Conversão de séries temporais em redes complexas

A motivação deste trabalho está nos bons resultados que o uso de abordagens de grafos e redes complexas tem demonstrado no estudo de sinais de epilepsia. Nesta seção é feita uma revisão bibliográfica sobre redes complexas e algumas de suas métricas.

2.3.1 Redes complexas

O famoso problema das setes pontes de Königsberg, provado por Leonhard Euler em 1735, marca a fundação da teoria dos grafos (NEWMAN, 2003). Com o passar do tempo essa teoria ganhou corpo e virou a base do estudo de redes. Na atualidade o grande acúmulo de informação trouxe à tona o fato de que as soluções dos métodos tradicionais da teoria dos grafos, que procuram por soluções ótimas, não são mais viáveis.

Começa então o movimento de estudo de redes de grande porte com um enfoque mais estatístico. Esse movimento dá o nome de seu objeto de estudo de redes complexas. Até então, a análise era feita partindo-se de cada indivíduo do sistema expandindo esse comportamento para se entender o sistema.

Contudo, sabe-se por exemplo, que mesmo entendendo o funcionamento dos neurônios não é possível explicar a consciência ou a memória. A chave para tal entendimento está na interação dos elementos, que antes era desconsiderada. A abordagem das redes complexas permite representar diferentes sistemas de forma unificada, o que possibilita encontrar características e propriedades comuns antes não percebidas.

Uma rede complexa, portanto, trata do estudo de sistemas modelados utilizando-se grafos nos quais cada nó representa um componente do sistema, e as arestas a interação entre os nós. A interação entre os nós gera um comportamento que não pode ser explicado pelo estudo do comportamento dos nós individualmente.

A seguir serão apresentados os conceitos utilizados no estudo de redes complexas.

Um grafo G é definido como um par (V, E) , onde V é um conjunto finito e não vazio composto por n vértices e E é um conjunto de m arestas que conectam vértices de tal forma que $E \subseteq \{(u, v) \mid u, v \in V\}$. Em um grafo não-direcionado, a existência de uma aresta (i, j) implica que (j, i) também exista. Ou seja, o relacionamento E é dito simétrico: $\forall (i, j) \in E \implies (j, i) \in E$. Em um grafo direcionado, a existência de uma aresta (i, j) não implica que (j, i) também exista. Nesse caso, a aresta direcionada (arco) (i, j) significa que o vértice i está conectado ao vértice j , mas a recíproca não é verdadeira, ou seja $(i, j) \neq (j, i)$.

Caminho em um grafo é um conjunto ordenado de vértices saindo de um vértice de origem e indo a um vértice de destino. Cada transição de um vértice a outro só é feita se existe uma aresta conectando esses dois vértices. O comprimento de um caminho é dado pelo número de arestas necessárias para fazer as transições de um vértice a outro.

O comprimento do menor caminho entre dois vértices i e j é definido como l_{ij} . E representa o caminho de menor comprimento entre os vértices i e j . Se i e j não estão conectados, ou seja, não é possível chegar em j a partir de i , l_{ij} é definido como infinito (BARRAT; BARTHELEMY; VESPIGNANI, 2008).

2.3.2 Métricas para caracterização de redes complexas

É convencional neste texto o uso do símbolo $\langle z \rangle$ para denotar o valor médio da variável z .

- **Grau de um vértice**

O grau k_i de um vértice i é definido como o número de arestas do grafo, incidentes ao vértice i (BARRAT; BARTHELEMY; VESPIGNANI, 2008).

- **Entropia**

Entropia é uma medida de incerteza de uma dada característica. Aplicada a grafos pode-se usar a entropia do grau, relacionando a incerteza geral do grau na dada rede. Para o cálculo da entropia é necessário a distribuição da característica de interesse. A entropia do grau é dada por:

$$H = - \sum_i P(k_i) \log(P(k_i)) \quad (2.6)$$

É definido que $0 \log(0) = 0$.

- **Coeficientes de aglomeração**

O conceito de aglomeração em grafos refere-se à tendência de redes reais formarem cliques na vizinhança de um dado vértice. Clique é um subgrafo totalmente conectado. Portanto, aglomeração implica que se um vértice i está conectado com um vértice j , e j está conectado à l , então existe uma probabilidade alta de i também estar conectado à l . O coeficiente de aglomeração local de um dado vértice i é dado por:

$$C(i) = 2e_i / (k_i(k_i - 1)) \quad (2.7)$$

e_i é o número de arestas entre os vizinhos de i

k_i é o grau do vértice i

A média do coeficiente de aglomeração local é:

$$\langle C \rangle = 1/N \times \sum_i C(i) \quad (2.8)$$

Outro coeficiente de aglomeração é dado pela fórmula da transitividade, sendo definido como:

$$C_{\Delta} = \frac{3 \times \text{número de triplas completamente conectadas}}{\text{número de triplas}} \quad (2.9)$$

Uma tripla é definida como um subgrafo contendo exatamente três nós ([BARRAT; BARTHELEMY; VESPIGNANI, 2008](#)).

- **Eficiência**

A medida da eficiência da rede mede o tráfego de informação, ou seja, o quão eficientemente a rede troca informação. A eficiência é definida como ([LATORA; MARCHIORI, 2001](#)):

$$E = 1 / (N(N - 1)) \times \sum_{ij} 1 / l_{ij} \quad (2.10)$$

N é o número de vértices da rede

l_{ij} é o tamanho do menor caminho entre i e j

- **Complexidade**

A complexidade de uma rede é uma medida tomada em relação ao grau, definida como a razão entre o segundo momento do grau pelo valor médio do grau ($\langle k^2 \rangle / \langle k \rangle$). E mede o quão conectada é a rede, sendo útil quando a natureza heterogênea da rede afeta os processos em equilíbrio e não-equilíbrio (BARRAT; BARTHELEMY; VESPIGNANI, 2008).

- **Dominância do ponto central**

O quanto uma rede se aproxima de uma estrela, maior a medida da intermediação (do inglês *betweenness*) de qualquer ponto da rede. Sendo 0 para grafos completos, e 1 para grafos em forma de roda, no qual há um vértice central que participa de todos os menores caminhos. A dominância do ponto central foi introduzida por (FREEMAN, 1977), e é definida da seguinte forma:

$$C_D = \sum_i \frac{B_{max} - B_i}{N}, \quad (2.11)$$

onde B_i é o grau de intermediação do nó i e B_{max} é o maior valor de intermediação do grafo.

- **Grau médio dos vizinhos (k_{nn})**

O grau médio dos vizinhos de um vértice, denotado por $k_{nn,i}$, é definido como:

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in v(i)} k_j, \quad (2.12)$$

onde $v(i)$ representa o conjunto dos vizinhos de i .

O interesse no cálculo de tal medida não está nela em si, mas sim em uma medida derivada dela, o grau médio dos vizinhos como função do grau:

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i/k_i=k} k_{nn,i}, \quad (2.13)$$

sendo N_k o número de vértices com grau k , e o somatório sobre os vértices com grau k .

Esta medida é interessante pois é menos afetada por flutuações estatísticas, já que faz a média dos graus médios dos vizinhos, do que a probabilidade condicional $P(k'|K)$, e portanto é uma boa candidata para quando os dados não são extremamente grandes ou precisos (BARRAT; BARTHELEMY; VESPIGNANI, 2008). Ela é conveniente para analisar a função de correlação do grau, $k_{nn}(k)$, para vértices de grau k .

Se o grau dos vértices vizinhos é não correlacionado, $k_{nn}(k)$ será constante. Na presença de correlação, $k_{nn}(k)$ será crescente ou decrescente. Caso crescente a correlação é dita assortativa, e se decrescente é dita disassortativa.

• Assortatividade

Correlações entre as medidas é uma maneira interessante de se entender a estrutura da rede. Um tipo muito recorrente de mistura de padrões referido como assortatividade aparece em ecologia, epidemiologia e ciências sociais, mede a tendência de nós conectarem-se com outros nós que apresentem propriedades similares (BARRAT; BARTHELEMY; VESPIGNANI, 2008). Uma maneira de medir-se a assortatividade é pela relação entre o grau médio dos vizinhos de um determinado vértice. O coeficiente de assortatividade de Pearson r , é definido em (NEWMAN, 2002) como:

$$r = \frac{\sum_e j_e k_e / E - [\sum_e (j_e + k_e) / (2E)]^2}{[\sum_e (j_e^2 + k_e^2) / (2E)] - [\sum_e (j_e + k_e) / (2E)]^2}, \quad (2.14)$$

onde j_e e k_e são os grau nas extremidades da aresta e , e E é o número total de arestas.

Várias medidas existem para caracterizar a centralidade de um nó da rede. As mais comuns são centralidade do grau, de aproximação (do inglês *closeness centrality*) e de intermediação (BARRAT; BARTHELEMY; VESPIGNANI, 2008).

Comumente, a centralidade de um vértice ou aresta é usada para medir sua importância, e tal medida depende das características ou propriedades específicas de interesse. Há muitas medidas de centralidade. Tipicamente arestas são caracterizadas pela sua centralidade de intermediação, enquanto que vértices pela centralidade do grau, centralidade de proximidade ou centralidade de intermediação do vértice (BARRAT; BARTHELEMY; VESPIGNANI, 2008). Vértices também têm definido centralidade por autovetor (do inglês *eigenvector centrality*), *PageRank* e *k-core*.

• Centralidade de intermediação

Esse medida de centralidade considera quantitativamente nós importantes para a conexão de regiões da rede. O conceito foi introduzido por Freeman (1977 apud NEWMAN, 2001), e é definido como o número de menores caminhos entre um par de vértices que passam por um dado vértice (BARRAT; BARTHELEMY; VESPIGNANI, 2008).

$$b_i = \sum_{h \neq j \neq i} \sigma_{hj}(i) / \sigma_{hj} \quad (2.15)$$

$\sigma_{hj}(i)$ é o número de menores caminhos entre h e j que passam por i

σ_{hj} é o número total de menores caminhos entre h e j

• Centralidade de autovetor

Pode ser interpretado como uma extensão natural da centralidade do grau. Quantifica a importância que um nó importante tem em seus vizinhos, ou seja, a importância de um nó

pode aumentar se ele estiver conectado com um nó importante. A centralidade de autovetor dá a cada vértice uma pontuação proporcional à soma das pontuações de seus vizinhos (NEWMAN, 2010).

A centralidade de autovetor foi proposta por Bonacich (1987):

$$x_i = k_1^{-1} \sum_j A_{ij} x_j, \quad (2.16)$$

com k_1 sendo o maior autovalor da matriz de adjacência A .

Uma característica interessante da centralidade de autovetor é poder ser alto ou por ter um vértice com muitos vizinhos, ou por tal vértice ter vizinhos importantes, ou ambos.

- **PageRank**

PageRank é um algoritmo baseado na ideia de que um *ranking* viável depende da estrutura topológica da rede, e é obtido pela simulação de uma navegação aleatória sobre o grafo da web (BARRAT; BARTHELEMY; VESPIGNANI, 2008). No contexto de centralidade, o *PageRank* dá uma medida da importância dos vértices.

Segundo o artigo original no qual ele foi proposto (BRIN; PAGE, 1998), o *PageRank* de um vértice i é definido por:

$$P_R(i) = \frac{q}{N} + (1 - q) \sum_j x_{ij} \frac{P_R(j)}{k_{out,j}}, \quad (2.17)$$

onde x_{ij} é um elemento da matriz de adjacência, $k_{out,j}$ é o grau de saída do vértice j , N é o número total de páginas no grafo da *world wide web* (número de vértices), e q é o chamado *damping factor*. O *damping factor* modela a probabilidade de um usuário parar uma navegação seguindo os *links* do *website* atual e começar uma nova navegação de um *website* totalmente novo, digitando um novo endereço.

- **Centralidade por proximidade**

Centralidade por proximidade mede a distância média de um dado vértice aos outros, sendo bem diferente das medidas de centralidade anteriores. A distância média entre vértices é calculada utilizando o conceito de caminho geodésico, o menor caminho entre dois vértices, o mesmo tipo de caminho utilizado na centralidade de intermediação.

Esta medida quantifica o quão fácil é chegar em outros vértices partindo de um vértice em particular, e é definida por:

$$C_i = \frac{n}{\sum_j d_{ij}}, \quad (2.18)$$

onde n é o número de vértices da rede, e d_{ij} é o tamanho do menor caminho entre i e j (NEWMAN, 2010).

- ***K-core***

Um *k-core* é definido como o maior subgrafo no qual cada vértice está conectado a pelo menos k outros vértices do subconjunto. Como é fácil encontrar todos os *k-core* de uma dada rede, esta medida tem interesse prático na análise da rede.

Um algoritmo simples, começa com a rede inteira e remove vértices com grau menor que k . Após recursivamente todos os vértices com grau menor que k tiverem sido removidos o subgrafo resultante é o *k-core* da rede original (NEWMAN, 2010).

Se um dado vértice pertencer ao *k-core* mas não ao $(k+1)$ -core então o *coreness*, ou o *k-core*, desse vértice é k . A medida de centralidade, derivada do *k-core*, que foi utilizada foi o *coreness* dos vértices. É suposto que quanto maior o *coreness* de um vértice maior sua centralidade, ou seja, sua importância.

2.3.3 Outros conceitos de redes complexas

Além das métricas descritas para caracterizar redes complexas, outros conceitos importantes também são utilizados na classificação de sinais eletrofisiológicos como o EEG. Como no contexto deste trabalho, as métricas são as características mais utilizadas, esses conceitos serão brevemente descritos pois são importantes na área de redes complexas. Newman (2010) trata desses conceitos com mais profundidade.

O primeiro desses conceitos é o de comunidade. Comunidade é um grupo de vértices densamente conectado, com poucas conexões entre grupos. Na detecção de estrutura de comunidades, assume-se que a rede divide-se naturalmente em subgrupos e a questão é achar tais grupos.

Por outro lado, *motifs* são padrões de interconexão de subgrafos cuja probabilidade de aparecer em um grafo aleatório equivalente é maior ou igual ao número de vezes que aparece na rede real, e menor que um valor de corte (e.g. 0,001) (MILO *et al.*, 2002). Eles tipicamente representam uma característica relativa à rede real.

A análise de propriedades estatísticas e topológicas de redes já produz uma rica quantidade de informações sobre os processos dinâmicos modelados por tais redes. Mas as vezes não se tem uma rede para analisar, nesses casos usa-se então modelos que representem o processo ou estrutura de interesse.

Os modelos permitem estudar os efeitos de determinadas características que uma rede real apresenta e extrapolar para um sistema maior. Isso considerando que as redes reais para serem analisadas foram amostradas e, em geral, não se amostra a rede inteira. Os modelos permitem analisar o efeito de determinada característica afetar algum processo dinâmico de interesse. Por exemplo, qual a influência da distribuição do grau na propagação de epidemias.

Existem vários modelos, cada um tenta suprir uma deficiência dos modelos anteriores.

- **Modelo Erdős-Rényi ou rede aleatória (ER):** Este é um modelo de grafo aleatório. Essa classe de modelos tem um conjunto de parâmetros com valores fixos mas a rede é aleatória em outro aspecto. No modelo ER, denotado por $G(n, p)$, é fixado o número de vértices, n , e a probabilidade de existência de uma aresta entre dois vértices quaisquer, p ;
- **Modelo Watts-Strogatz ou rede pequeno mundo (WS):** Aqui o objetivo é gerar redes com alto coeficiente de aglomeração. Para tanto esse modelo faz com que todo vértice esteja conectado com m vizinhos de cada lado. Desta maneira a rede gerada terá alta transitividade, e esse valor não diminui conforme a rede aumenta;
- **Modelo Barabási-Albert ou rede livre de escala (BA):** Neste modelo, vértices são adicionados um a um, a uma rede crescente, e cada vértice conecta-se a um conjunto adequado de vértices previamente existente. Esse modelo é baseado no crescimento e na ligação preferencial. No processo de crescimento, a cada passo, um vértice é adicionado com m ($m \leq m_0$) arestas que conectam m vértices presentes na rede.

2.3.4 Métodos de conversão de séries temporais em redes complexas

Nesta seção serão relacionados os trabalhos voltados às técnicas de mapeamento de séries temporais para redes complexas.

[Donner et al. \(2011\)](#) fazem um estudo focado em redes de recorrência, mas também descrevem outras maneiras de mapear séries temporais em redes complexas. Nesse trabalho os autores dividem esses métodos de transformação em três classes: redes de proximidade, grafos de visibilidade e redes de transição.

- **Redes de proximidade:** segmentos da série temporal são representados por vértices e medidas de similaridade definem as conexões entre os vértices. Devido aos vários métodos de calcular a similaridade, essa abordagem é subdividida em três: redes cíclicas, redes de correlação e redes de recorrência. Cada uma dessas será apresentada a seguir.

- **redes cíclicas:** [Zhang e Small \(2006\)](#) desenvolveram um trabalho no qual séries pseudoperiódicas são mapeadas em um grafo. O ciclo é considerado como uma subsequência entre dois mínimos, ou máximos, locais. Foram estudados séries de batimentos cardíacos e fala humana. A série é dividida em m ciclos $\{C_1, C_2, C_3, \dots, C_m\}$ cada um representando um vértice, e uma aresta entre eles é criada baseado na correlação ρ entre os ciclos da série. Essa correlação é a similaridade, medindo o quão semelhantes são dois ciclos.

É definido um limiar D tal que se $\rho_{i,j} \geq D$, uma aresta é criada entre os vértices C_i e C_j . Após a construção da rede, medidas de caracterização como distribuição do grau, coeficiente de aglomeração e média dos menores caminhos são calculadas para as

séries testadas. Séries periódicas ruidosas produzem redes aleatórias, séries caóticas geram redes com propriedades de pequeno mundo e livre de escala.

Um trabalho posterior (ZHANG *et al.*, 2008) utilizou outras medidas como assortatividade, centralidade de intermediação e *motifs* para traçar um paralelo entre as dinâmicas das séries e as medidas estatísticas da rede, possibilitando a diferenciação de séries temporais de dinâmicas diferentes com essas medidas;

- **redes de correlação:** Yang e Yang (2008) construíram uma rede complexa a partir da matriz de correlação de uma série temporal. É calculada a correlação $r_{i,j}$ entre pares de subsequências S_i e S_j de comprimento l de uma dada série. Cada subsequência é representada por um vértice, e se $r_{i,j} \geq r_C$, onde $r_C \in [0, 1]$ é um limiar definido, uma aresta entre os vértices i e j é criada. Para r_C grande poucas conexões são formadas e mais ruído é eliminado da série, mas arestas importantes também podem ser removidas.

Yang e Yang (2008) focam nesses parâmetros, l e r_C , e também analisam a distribuição do grau dessa rede e seu ajuste à lei de potência, reiterando o potencial do uso de redes complexas para análise de séries temporais.

- **redes de recorrência:** Um espaço de fase, ou espaço de estado, de uma série temporal é um espaço matemático que representa todos os possíveis estados que um sistema dinâmico que representa uma série temporal pode ter. Tipicamente, sistemas dinâmicos são caracterizados por caminhos ou trajetórias nesse espaço. O chamado gráfico de recorrência (do inglês *recurrence plot*), ou matriz de recorrência, é dado pela expressão:

$$R_{i,j} = \Theta(\varepsilon - d(\vec{x}_i, \vec{x}_j)), \quad (2.19)$$

onde $\Theta(\cdot)$ é a função *Heaviside*, $\vec{x}_i$ e $\vec{x}_j$ são vetores que representam pontos do espaço de fase, $d(\cdot)$ é uma função de distância que mede a proximidade entre dois pontos no espaço de fase, tipicamente a distância euclidiana, e ε (*epsilon*) é um limiar de proximidade.

A matriz R determina a proximidade de pontos no espaço de fase, ela é uma matriz binária, e é utilizada para a construção da rede complexa. Cada ponto do espaço de fase é representado por um vértice e a recorrência entre eles indica uma aresta ligando-os (MARWAN *et al.*, 2009). A recorrência é dada pela matriz de recorrência. O coeficiente de aglomeração e a média dos menores caminhos são as medidas utilizadas para caracterizar a dinâmica da rede no trabalho de Yang e Yang (2008).

Diferente de Marwan *et al.* (2009), que utilizam um limiar ε para conectar vértices, outras abordagens constroem as arestas baseado nos k vizinhos mais próximos e k vizinhos mais próximos adaptativos (SHIMADA; KIMURA; IKEGUCHI, 2008; XU; ZHANG; SMALL, 2008; DONNER *et al.*, 2011). A abordagem do k vizinhos mais

próximos adaptativos realiza um processo diferente do k vizinhos mais próximo de forma a gerar um grafo não direcionado e com grau médio $\langle k \rangle = 2E_0$, onde E_0 é o equivalente ao k do k vizinhos mais próximos.

- **Grafos de visibilidade:** Para construir o grafo de visibilidade as amostras da série temporal representam os vértices, e são criadas arestas para todos os vértices entre os quais é possível traçar uma reta, e essa reta fica acima dos vértices intermediários ao início e fim da reta. [Lacasa et al. \(2008\)](#) exemplificam plotando a série como um gráfico de barras (figura 4), o valor de cada observação da série dita a altura da barra. Uma aresta entre i e j é criada se é possível traçar uma reta da barra do vértice i até a barra do vértice j sem cruzar por nenhuma outra barra. Uma vantagem é que essa rede formada é invariante a transformações da série.

Figura 4 – Exemplo de um grafo de visibilidade. Adaptado de ([LACASA et al., 2008](#)).

Quando séries periódicas são convertidas, a rede resultante é regular; para séries aleatórias redes aleatórias são criadas, e séries fractais são mapeadas para redes livre de escala. Algumas aplicações dessa abordagem estão relatadas em [Elsner, Jagger e Fogarty \(2009\)](#) para detectar anos anormais quanto à contagem de furacões; [Shao \(2010\)](#) aplica em uma série de batimentos cardíacos, e identifica pacientes com insuficiência cardíaca pela assortatividade; [Nuñez et al. \(2012a\)](#) detectam a periodicidade de séries temporais usando grafos de visibilidade, e por fim [Nuñez et al. \(2012b\)](#) fazem uma revisão de grafos de visibilidade;

- **Redes de transição:** Esta abordagem produz um grafo ponderado completo, em que os vértices do grafo representarão estados discretizados do intervalo de valores da série, chamados de classes, e as arestas são dadas pela probabilidade de transição de um estado a outro. O peso da aresta é a probabilidade condicional $P(X_{i+1} \in L_i | X_i \in L_j)$, ou seja, dado

que a observação X_i da série pertença à classe L_j , qual a probabilidade do estado X_{i+1} pertencer a classe L_i .

Esse grafo pode ser convertido em um grafo não ponderado utilizando-se um limiar de conexão p , abaixo do qual as arestas são removidas. Diferentemente das outras abordagens, a topologia da rede depende apenas do método de segmentação escolhido. Alguns dos principais trabalhos que utilizam essa abordagem são [Nicolis, Cantú e Nicolis \(2005\)](#), [Li e Wang \(2007\)](#), [Li et al. \(2007\)](#).

Há ainda duas outras maneiras de converter as séries temporais não relatadas nos trabalhos acima. O mapeamento de processos estocásticos a redes complexas e uma modificação do grafo de visibilidade, o grafo de visibilidade binário de *Markov*.

[Shirazi et al. \(2009\)](#) introduzem uma maneira de mapear processos estocásticos em redes complexas. Foram também estudadas algumas medidas de caracterização de redes complexas, o resultado mais interessante foi a reconstrução com alta precisão da série original através de uma caminhada aleatória na rede.

[Ahadpour, Sadra e ArastehFard \(2014\)](#) analisam alguns aspectos topológicos como grau, coeficiente de aglomeração e caminho médio, do grafo de visibilidade binário de *Markov*. Nessa construção, a série é discretizada e se o valor da série cresce ele vira o valor 1 se ele diminui ele vira o valor 0, e a partir dessa nova série é criado o grafo de visibilidade. Essa abordagem mostrou-se simples para a discriminação de sistemas dependentes ou independentes.

2.3.5 Análise de séries temporais por meio de redes complexas

Nesta seção serão apresentados alguns dos trabalhos presentes na literatura que usam redes complexas para a análise de séries temporais, a maioria dos trabalhos são focados na caracterização de sinais neurais, mas não exclusivamente. Conhecendo-se as medidas de caracterização de redes complexas, tem-se condição de estudar como os grafos mapeados a partir de séries temporais são úteis para a análise de tais séries.

[Reijneveld et al. \(2007\)](#) enfatizam a correlação entre as propriedades estruturais das redes e as dinâmicas presentes nas mesmas. É mostrado que as redes funcionais e anatômicas (estruturais), em humanos saudáveis, têm muitas características de redes pequeno mundo, mas apenas um pouco de características de redes livre de escala. Os autores especulam que as redes neurais desenvolvem-se em direção a um regime com um limiar crítico entre uma sincronização local e global. E crises resultam de uma patologia que levam o sistema além desse limiar.

[Schindler et al. \(2008\)](#) utilizam o coeficiente de aglomeração e o caminho característico como indicativo do movimento de uma rede aleatória para uma rede mais regular e então de volta. Esse movimento é devido a uma crise epilética sofrida pelo paciente em questão. O trabalho destaca a alta relevância do uso de redes complexas como meio de aumentar o entendimento

da complicada dinâmica das crises epiléticas. Para criar a rede complexa, os autores utilizaram a rede de recorrência da série em questão, e valores da matriz acima de determinado limiar significava que os nós, representando os índices da matriz, deveriam compartilhar uma conexão. O trabalho mostrou que independentemente de muitos fatores que tipicamente influenciam muito a análise, como o arranjo dos nós, a não uniformidade dos sensores, o foco epilético entre outros, a presente análise sempre caracterizou o movimento da rede para uma rede mais regular e depois de volta para uma rede mais aleatória.

[Bullmore e Sporns \(2009\)](#) fazem uma retomada do uso de redes complexas para análise de sinais neurais. Ressaltam características estruturais e funcionais do cérebro quando representadas em redes complexas, tal como topologia seguindo pequeno mundo, *hubs* conectados e modularidade. Por fim destaca desafios técnicos e questões chaves que precisarão ser trabalhados. Como o entendimento da relação da rede funcional com seu substrato estrutural, ou seja, como a topologia de redes funcionais mudam com o tempo.

[Horstmann et al. \(2010\)](#) mostram o uso de redes complexas para análise de epilepsia, nesse caso focado em redes funcionais. Analisaram a rede funcional de pacientes com epilepsia tratados com a rede de pacientes saudáveis. Utilizaram o coeficiente de aglomeração e o caminho médio como medidas de caracterização das redes.

O trabalho descrito nesta dissertação desenvolve um estudo utilizando o ferramental de redes complexas para a análise de sinais epiléticos, uma abordagem parecida com a de [Zhu \(2012\)](#) embora esta última seja menos completa. O trabalho citado emprega o algoritmo de conversão de séries temporais em redes pelo grafo de visibilidade. Nele, o sinal de EEG de um único canal foi mapeado no seu respectivo grafo de visibilidade, extraíndo-se a distribuição do grau e grau médio. Mostrou-se que o grau médio de pacientes com epilepsia é maior do que o de pacientes saudáveis. Com isso mostrou-se que o grafo de visibilidade tem alta acurácia na classificação de três tipos de sinais; EEG de paciente saudável, EEG de paciente com epilepsia em uma crise e EEG de paciente com epilepsia no período entre crises. O autor relata que embora trabalhos utilizando teoria dos grafos para análise de EEG epilético tenham sido feitos, nenhum utilizando grafo de visibilidade foi empregado até a divulgação desse trabalho, e que a característica de lei de potência do grafo de visibilidade pode ser usada para discriminação de EEG ictal e saudável.

[Subramaniyam e Hyttinen \(2013\)](#) também fazem análise do EEG de epilepsia utilizando redes complexas. Esse trabalho gera uma rede de recorrência do sinal para quantificar a dinâmica não linear dos atratores saudáveis, interictal e ictal. Foi computado o coeficiente de aglomeração, o comprimento dos caminhos, como meio de caracterizar essas redes. Mostrou-se que sinais interictais são caracterizados por atratores caóticos, com redes apresentando propriedades de pequeno mundo, como alto coeficiente de aglomeração e baixo caminho médio, enquanto que os sinais ictais são caracterizados por atratores quase periódicos com altos coeficientes de aglomeração e caminho médio. Uma vantagem da abordagem de redes de recorrência é que não foi feita nenhuma suposição quanto à estacionariedade do sinal, nem quanto ao seu comprimento,

e essa abordagem não é sensível a parâmetros utilizados na construção do espaço de fase.

Lang *et al.* (2013) realizam mais um trabalho que utiliza redes de recorrência para analisar EEG de pacientes saudáveis e com epilepsia. Os resultados são parecidos com os de Subramaniyam e Hyttinen (2013), mas também calculou-se a assortatividade, e os valores aumentam de cérebros normais para os com epilepsia. A assortatividade pode, portanto, ser usada para diferenciar esses estados.

Como a falta de simetria em relação ao tempo é uma característica importante para alguns tipos de séries temporais não lineares, Donges, Donner e Kurths (2013) utilizaram testes estatísticos com medidas de redes complexas para verificar a presença ou ausência de simetria nas séries temporais. Os autores analisaram o grau e o coeficiente de aglomeração para tal atividade. Esse trabalho mostra mais um uso de redes complexas para análise de características de sistemas não lineares. Um detalhe importante é que não se explorou informações temporais para a construção dos grafos, e outros métodos para construção de redes que não possuem essa característica, como as redes de recorrência, não podem ser utilizados para o mesmo propósito.

Straaten e Stam (2013) discutem aspectos técnicos do uso de redes para análise de sinais neurais. Um desenvolvimento importante que ainda precisa ser feito, segundo o autor, é a criação de um modelo que capture as nuances do cérebro melhor que os atuais.

Zhu, Li e Wen (2014) propõem um grafo de visibilidade ponderado rápido (FWHV) para identificar crises em EEG. Ele foi avaliado com outras abordagens, transformada rápida de Fourier (FFT) e entropia de amostras (SampEn), e se mostrou robusto ao ruído. Teve custo computacional mais baixo que os outros métodos comparados, e com uma acurácia na classificação de crises de 100%. O autor relacionou trabalhos passados que utilizaram outras métricas e abordagens na classificação de sinais epiléticos, e escolheu a FFT devido a seu baixo custo computacional comparado com as outras métricas analisadas. O trabalho menciona que mais estudo é necessário quanto ao tamanho do sinal para esse tipo de abordagem, já que o tamanho influenciou no desempenho. Em relação aos outros métodos, o FWHV é uma ferramenta eficiente para o processamento de EEG em tempo real dados os parâmetros típicos utilizados, como tamanho da época.

Aggarwal e Subbian (2014) fazem uma revisão de técnicas aplicadas a redes dinâmicas, ou seja, que evoluem com o tempo. Como a proposta desse trabalho pode ser considerada como a de uma rede que evolui, esse trabalho relaciona outros trabalhos e técnicas que podem vir a ser utilizados.

Para Bhaduri e Ghosh (2014), as técnicas usuais de análise de dados de EEG não capturam algumas propriedades importantes para a avaliação quantitativa do progresso da disfunção do cérebro humano. O grafo de visibilidade empregado nesse trabalho consegue capturar a multifractalidade acuradamente, além de resultados passados terem mostrado que essa técnica provê resultados confiáveis para dados comparativamente curtos. A potência, o

coeficiente da lei de potência, ligada à livre escala do grafo conseguiu efetivamente medir a fractalidade do sinal. Essa técnica mostrou a primeira técnica quantitativa para medir o grau da fractalidade, de que alguma forma mede a complexidade do sinal. Essa técnica, que se mostrou robusta, também foi capaz de distinguir a doença com respeito a sua severidade.

Yu *et al.* (2016) faz a comparação dos coeficientes de aglomeração com o espectro de potência de EEG de pacientes saudáveis e com epilepsia com o espectro para determinar qual é o mais discriminativo. Os coeficientes de aglomeração são extraídos de uma rede k vizinhos mais próximos adaptativos construída do EEG. Ele conclui que o espectro de potência torna essa discriminação mais fácil, mas o coeficiente de aglomeração também é capaz de distinguir entre os sinais.

Bhaduri e Ghosh (2015), Wang *et al.* (2016), Supriya *et al.* (2016), Liu *et al.* (2016) e Hao, Chen e Zhao (2016) fazem uso do grafo de visibilidade, ou alguma versão modificada do mesmo, para analisar sinais de EEG de epilepsia. Em geral esses trabalhos convertem o sinal numa rede complexa, em seguida extraem uma ou várias das medidas grau e peso das arestas, propriedades de comunidade, centralidade do grau, e utilizam essas medidas como entrada em algum classificador que então rotula o sinal. Os classificadores utilizados foram k vizinhos mais próximos (Knn), SVM e *Linear Discriminant Analysis* (LDA).

2.4 Considerações finais

Apresentada a revisão bibliográfica, pretende-se seguir a seguinte linha: fazer a classificação dos sinais de EEG de pacientes com epilepsia através da conversão do sinal em redes complexas utilizando as várias abordagens descritas. O sinal é analisado em janelas. Em cada grafo são extraídas várias métricas e com isso um novo conjunto de dados no formato atributo-valor é criado. Esse conjunto de dados é utilizado no treino de classificadores que são utilizados para detectar quando ocorre uma crise. O método de trabalho é descrito em detalhes no próximo capítulo.

MATERIAIS E MÉTODOS

Na seção de materiais são descritas as bases de dados utilizadas.

A seção métodos contém uma descrição do fluxo de trabalho para atingir os objetivos propostos, bem como as métricas de avaliação e comparação de resultados.

3.1 Materiais

Os materiais considerados nesse trabalho são as bases de dados que foram usadas.

A primeira base de dados aqui especificada é a mais utilizada nos trabalhos pesquisados, e foi descrita por Andrzejak *et al.* (2001)¹. São 5 conjuntos (denotados de A a E), cada um contendo 100 segmentos de EEG com duração de 23,6 segundos de um canal. Esses segmentos foram selecionados e retirados de uma gravação contínua multicanal de EEG após inspeção visual para a não inclusão de artefatos. Os segmentos selecionados foram os que não rejeitaram a hipótese de que esses segmentos tinham as mesmas propriedades (determinismo e estruturas de baixa dimensão) de um processo estocástico linear *gaussiano* e estacionário, sendo usada a estacionariedade fraca para evitar rejeições da hipótese nula devido apenas à não estacionariedade.

Os conjunto A e B consistem de segmentos de gravação de EEG do couro cabeludo realizadas em 5 pacientes saudáveis usando eletrodos posicionados segundo o padrão internacional 10-20. Os voluntários estavam relaxados, acordados e com os olhos abertos no conjunto A e com olhos fechados no conjunto B. Os conjuntos C, D e E são de arquivos de diagnósticos pré-cirúrgicos.

Nesse estudo, 5 pacientes foram selecionados, todos com completo controle das crises após ressecção de uma das formações no hipocampo, a qual foi diagnosticada como a zona epileptogênica. Os segmentos do conjunto D foram gravados de uma zona epileptogênica, e os

¹ Disponível em: <http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3>

do segmento C da formação do hipocampo no hemisfério oposto do cérebro. Os conjuntos C e D contêm apenas atividade medida durante períodos livres de crise, o conjunto E contém apenas período de crise. Nesse caso, os segmentos foram selecionados de todos os locais exibindo atividade ictal.

Todos os sinais de EEG foram gravados com o mesmo sistema de amplificação de 128 canais, usando uma média comum de referência, mas omitindo eletrodos contendo atividade patológica (conjuntos C, D e E) ou artefatos muito pronunciados (conjuntos A e B). Uma conversão analógico-digital de 12 bits foi usada, e os dados foram salvos continuamente num disco rígido no equipamento de aquisição com uma amostragem de 173,61 Hz. Um filtro passa-banda foi aplicado com banda passante de 0,53 a 40 Hz.

A nomenclatura dos conjuntos A, B, C, D e E é a definida pelo trabalho que descreve o conjunto de dados ([ANDRZEJAK *et al.*, 2001](#)), mas os arquivos com os dados do conjunto de dados utilizam outra, O, Z, N, F e S, respectivamente. Esta última será a nomenclatura utilizada nesta dissertação.

A figura 5 mostra um exemplo de cada um dos 5 conjuntos.

Os dois conjuntos de dados sintéticos descritos abaixo foram utilizados apenas na análise de estacionariedade para ter um melhor controle e entendimento sobre a operação do método proposto e para explorar melhor as perspectivas estocástica e determinística, assim como o método de análise que se propõe.

O foco do primeiro conjunto é a análise estocástica e tem dois tipos de segmentos, cada um de uma função de probabilidade diferente. A distribuição normal padrão, média zero e variância um, combinada com outra distribuição normal com média zero e variância dois.

Uma série é formada pela combinação de dois tipos de partes, cada uma com 4.096 pontos amostrados das distribuições descritas acima, totalizando 81.920 pontos (10 partes). A figura 6 ilustra um segmento dessa série.

A perspectiva determinística foi avaliada usando dados do sistema de *Lorenz* ([LORENZ, 1963](#)). Ele é composto por três equações com três variáveis, x , y , e z . Para simular a não estacionariedade, os parâmetros do sistema foram configurados para assumir valores dentre dois conjuntos. Os conjuntos de parâmetros são $\sigma = 10$, $\rho = 28$ e $\beta = 2,667$ para a primeira parte do sinal, e $\sigma = 10$, $\rho = 28$ e $\beta = 3,5$ para a segunda. Abaixo, é possível ver como cada parâmetro influencia o sistema (Eqs. 3.1, 3.2 e 3.3).

$$\frac{dx}{dt} = \sigma(y - x) \quad (3.1)$$

$$\frac{dy}{dt} = x(\rho - z) - y \quad (3.2)$$

Figura 5 – Séries da Universidade de Bonn. A - Paciente saudável com olhos abertos. B - Sinal de EEG durante uma crise.

$$\frac{dz}{dt} = xy - \beta z \quad (3.3)$$

O passo temporal usado na iteração das equações, o valor de dt , foi 0,01, e a posição inicial do sistema foi $(x; y; z) = (0,00; 1,00; 1,05)$.

O primeiro conjunto de parâmetros é usado e então espera-se o sistema alcançar um comportamento assintótico; para essa espera usou-se 4.096 pontos, então o segundo conjunto de parâmetros é usado, e novamente depois do sistema manter o regime assintótico por um tempo os parâmetros são mudados para o primeiro conjunto, e assim por diante. A figura 7

Figura 6 – Trecho do conjunto da série estocástica usada.

mostra a série com dois pedaços do sinais, um para cada conjunto de parâmetros.

Como esta série é determinística, apenas três mudanças de parâmetros foram usadas.

O conjunto da Universidade de Bonn também foi utilizado na análise de estacionariedade. Mas para avaliar a identificação de mudança de comportamento, a entrada do método foi duas séries de conjuntos diferentes unidas formando uma com dois comportamentos diferentes. Foram testadas 30 séries dos conjuntos A (pacientes saudáveis com olhos abertos) e E (durante uma crise) unidas, totalizando 8.194 pontos cada nova série.

3.2 Métodos

Os métodos utilizados neste trabalho foram divididos em dois grupos: análise de estacionariedade e classificação de crises epiléticas.

Primeiro, é descrita a análise de estacionariedade, em seguida como é feita a classificação das crises.

Todo o desenvolvimento do presente trabalho foi feito na linguagem *python* utilizando bibliotecas de código livre, e quando o método utilizado não estava disponível o mesmo foi implementado. Todos os códigos estão sendo integrados numa biblioteca e a mesma está disponibilizada no seguinte repositório aberto <<https://github.com/danielcestari/ts2cn>>.

3.2.1 Estacionariedade

Como descrito anteriormente, é importante checar se os dados sendo analisados são estacionários. Aqui é descrita uma abordagem estocástica e determinística para a análise de

Figura 7 – Trecho do sinal determinístico (sistema de *Lorenz*) usado. A - Atrator do primeiro conjunto de parâmetros. B - Atrator do segundo conjunto de parâmetros. C - Série os dois conjuntos de parâmetros repetidos suas vezes.

estacionariedade; ambas as abordagens são relevantes e podem ser vistas como complementares, dependendo da perspectiva sendo analisada, uma pode ser mais adequada que a outra.

Além do objetivo em si do estudo da estacionariedade do sinal, outro resultado útil dessa análise é uma possível restrição nos valores do intervalo para os parâmetros utilizados posteriormente no estudo das séries temporais.

Abaixo são descritos os métodos propostos para a análise de estacionariedade de uma perspectiva estocástica e outra determinística seguindo o especificado no trabalho de [Cestari e Rosa \(2017\)](#).

3.2.1.1 Estocástica

A maneira de verificar a estacionariedade é checar se a distribuição da variável aleatória muda com o tempo. Uma maneira direta e muito difundida para verificar se houve mudança, é a aplicação de um teste de aderência, ou *goodness of fit test*. Existem vários destes testes, o que é usado neste trabalho é baseado no teste χ^2 . Este teste foi escolhido e recomendado por alguns trabalhos para testar a estacionariedade ([ZOLDI; KRYSTAL; GREENSIDE, 2000](#); [WITT; KURTHS; PIKOVSKY, 1998](#); [DIKANEV et al., 2005](#)). Outro teste disponível é o teste *Kolmogorov-Smirnov* de duas amostras, mas ele é mais sensível próximo ao centro da distribuição do que nas caudas. No contexto de EEG de epilepsia, isto pode ser um problema já que a presença de *spikes* é representada nas caudas.

Em cada configuração, o teste de hipótese utilizando o histograma de duas janelas consecutivas e um nível de significância é executado. Portanto, cada série temporal é avaliada para vários tamanhos de janelas e níveis de significância. O número de intervalos (do inglês *bins*) é passado como parâmetro.

$$\chi^2 = \sum_{i=1}^{N-1} \frac{(curr_i - next_i)^2}{curr_i + next_i} \quad (3.4)$$

O teste de hipótese é efetuado com a estatística χ^2 (Eq. 3.4) em que N é o número de intervalos, $curr_i$ e $next_i$ são as densidades do i -ésimo intervalo das janelas atual e próxima, no mecanismo da janela deslizante ([PRESS et al., 1996](#)). Por esta razão, ambos os histogramas precisam ter o mesmo suporte, ou seja, os mesmos intervalos, e o grau de liberdade usado no teste é $N - 1$. Este teste é executado para todas as janelas consecutivas da série.

Como resultado da abordagem, cada série temporal produz uma lista de valores p. Para um dado nível de significância, esta lista é convertida numa sequência de valores booleanos indicando se as janelas tiveram sua hipótese nula rejeitada ou não.

Para uma visão geral do desempenho na série temporal inteira, uma medida que agrega esta lista de booleanos é computada. Primeiramente, um rótulo é atribuído a cada janela, sendo este, o rótulo da amostra mais frequente dentro da janela. Com os rótulos das janelas e a lista

de valores p , a medida de desempenho é obtida penalizando quando janelas adjacentes têm o mesmo rótulo mas o teste rejeita a hipótese nula, ou quando os rótulos mudam mas o teste falha em rejeitar a hipótese nula. A medida é normalizada entre zero e um, significando a proporção das janelas cujo resultado do teste e os rótulos estão de acordo, ou seja, rejeita a hipótese nula quando há mudança de rótulos, e falha em rejeitá-la quando os rótulos não mudam.

Abaixo são listados os parâmetros e seus valores experimentados:

- **Tamanho de janela:** Os valores de tamanho de janelas utilizados na análise estocástica foram 128, 256, 512, 1.024, 2.048, 4.096 pontos. Esses valores foram escolhidos pois não há uma maneira ótima para a escolha do tamanho da janela, e esses valores cobrem tamanhos possíveis utilizados na literatura, além do fato de uma janela ter o dobro do tamanho da anterior minimizando a quantidade de tamanhos de janelas a ser testado;
- **Nível de significância (α):** Os níveis de significância utilizados nos testes de hipótese foram 0,01; 0,05; 0,1; 0,2; 0,3. Os níveis de significância 0,01; 0,05; e 0,1 são os tipicamente utilizados, e os valores 0,2 e 0,3 foram adicionados na análise para ir um pouco além do comumente utilizado.;
- **Número de intervalos (do inglês *bins*):** O valor do número de intervalos utilizados na construção dos histogramas foi o maior entre 20 ou o tamanho da janela dividido por 40. A quantidade de intervalos utilizada foi pensada para ser uma quantidade razoável, mas tentando evitar o problema de ter poucos pontos por intervalo.

3.2.1.2 Determinística

A análise de estacionariedade proposta aqui e chamada de determinística, segue os procedimentos detalhados por [Manuca e Savit \(1996\)](#).

Como foi anteriormente afirmado, determinar a estacionariedade significa determinar se o sistema dinâmico que descreve o fenômeno sendo analisado é autônomo ou não. A fim de se descobrir a dependência temporal, pode-se simplesmente analisar como algumas medidas que caracterizam o sistema evoluem com o tempo.

A medida usada para este fim é a proximidade dinâmica simétrica (Eq. 3.6). Ela é a versão simétrica da proximidade dinâmica (Eq. 3.5), que mede a proximidade, em termos de sistemas dinâmicos, entre duas janelas da série temporal sendo analisada. Ambas foram definidas por [Manuca e Savit \(1996\)](#). Elas se baseiam em outras medidas calculadas do espaço de fase da série temporal, cada medida captura características distintas do sistema. O uso do espaço de fase implica que o sistema é modelado como um sistema dinâmico. Analisando como a proximidade dinâmica evolui, é possível inferir se o sistema é autônomo ou não.

$$d^{(m)}(x \leftarrow y) = \sqrt{|R_m(x, y) - R_m(y, x)|} \quad (3.5)$$

$$d^{(m)}(x, y) = \max(d^{(m)}(x \leftarrow y), d^{(m)}(y \leftarrow x)) \quad (3.6)$$

O espaço de fase é uma maneira de representar um sistema dinâmico, e para obter tal representação é necessário conhecer o sistema de equações que descreve o sistema, ou seja, é uma representação teórica. Como geralmente os dados utilizados na análise são amostrados do sistema dinâmico representado pelo sistema de equações ou pelo espaço de fase, é possível reconstruir, dadas algumas limitações relacionadas à qualidade dos dados, o espaço de fase do sistema.

A base teórica que estabelece a possibilidade de reconstrução é dado por [Takens \(1981\)](#). Ele estabelece que apenas um sinal é necessário para reconstruir o espaço de fase de um sistema com $2n + 1$ equações. Este é um resultado importante, e a maneira de se utilizar esse sinal para reconstrução do espaço de fase é transformar esse sinal na chamada configuração de atraso de tempo (do inglês *time-delay*). Essa configuração transforma um sinal genérico $y(t)$ num conjunto de pontos n -dimensionais $[y(t), y(t - \tau), y(t - 2\tau), \dots, y(t - 2n\tau)]$. Essa reconstrução necessita de dois parâmetros, o atraso τ que cada componente sofre e o número de dimensões n deste espaço.

De acordo com [Fraser e Swinney \(1986\)](#) a melhor maneira de se calcular o atraso é pelo valor do primeiro mínimo da informação mútua da série. Como os dados utilizados no trabalho são ruidosos, considerou-se um parâmetro de erro para o cálculo do mínimo da informação mútua. Esse erro é calculado como uma porcentagem ao intervalo de valores que a informação mútua atinge para a dada série cujo espaço de fase está sendo reconstruído. Para a análise de estacionariedade considerou-se um erro de 2% para o cálculo do mínimo da informação mútua.

Outro parâmetro necessário para o cômputo da informação mútua é o número de intervalos (do inglês *bins*) utilizados. Utilizou-se o valor fixo de 10 intervalos para todas as análises e para todos os tamanhos de janelas. Esse parâmetro é necessário pois a informação mútua é calculada utilizando a distribuição dos dados, e nesse caso é preciso informar o número de intervalos na construção do histograma.

A dimensão do espaço de fase reconstruído, também chamada de dimensão de embutimento ou imersão, é um parâmetro dessa análise, e posteriormente são discutidos os valores utilizados nesse parâmetro.

Os componentes da proximidade dinâmica são explicados a seguir.

Como na abordagem estocástica, uma janela deslizante é usada. Cada janela, um segmento da série temporal, tem seu espaço de fase reconstruído, então é possível calcular a proximidade dinâmica para caracterizar a dinâmica das janelas. Tendo o espaço de fase, a

correlação integral de Grassberger-Procaccia (Eq. 3.7) é calculada.

$$C_m = P(\|x^{(m)}(i) - x^{(m)}(j)\| < \varepsilon) \\ = \frac{1}{N_{pairs}} \sum_{i \neq j} \theta(\varepsilon - \|x^{(m)}(i) - x^{(m)}(j)\|) \quad (3.7)$$

Essa medida representa a probabilidade de dois pontos, $x^{(m)}(i)$ e $x^{(m)}(j)$, estarem próximos entre si no espaço de fase de dimensão m (GRASSBERGER; PROCACCIA, 1984; MANUCA; SAVIT, 1996). Toda vez que o símbolo θ aparecer, ele estará representando a função *Heaviside*, ou função degrau unitário; esta função retorna o valor zero se seu argumento é negativo, e retorna um caso contrário.

Em seguida, duas outras medidas são calculadas.

$$S_m = \frac{C_{m+1}}{C_m} \quad (3.8)$$

A primeira é a previsibilidade (do inglês *predictability*, Eq. 3.8), introduzida por Savit e Green (1991). Ela representa a probabilidade condicional de pontos próximos entre si no espaço de fase reconstruído com dimensão m , estarem próximos entre si no espaço de fase reconstruído com dimensão $m + 1$. Ela representa a habilidade de prever o próximo ponto da série se as m dimensões passadas são conhecidas (SAVIT; GREEN, 1991).

$$R_m = \frac{S_m}{S_{m-1}} \quad (3.9)$$

A segunda é a taxa de previsibilidade (do inglês *ratio of predictabilities*, Eq. 3.9), ela mede a melhora na previsibilidade devido à inclusão da informação contida na m -ésima dimensão (MANUCA; SAVIT, 1996).

Com essas medidas é possível retratar a proximidade dinâmica de todas as janelas. Como a dinâmica pode ter um amplo intervalo ao longo do tempo, uma versão normalizada da proximidade dinâmica simétrica (Eq. 3.10) é usada na análise.

$$M_{ij} = \sum_{k=1}^n \theta(|d_{ik} - d_{jk}| - \sqrt{2}) \quad (3.10)$$

Existe um tamanho de janela para o qual a força que move o sistema de seu ponto de operação, que é uma abstração da mudança dos valores dos parâmetros, será virtualmente constante na janela, mas diferente entre janelas (MANUCA; SAVIT, 1996). Aumentando o tamanho da janela, finalmente, o sistema será caracterizado como estacionário. Ou então, grupos de janelas podem ser vistos como tendo a mesma dinâmica, ou seja, estacionário naquele

agrupamento. Se o sistema é de fato estacionário, tal caracterização sempre será alcançada contanto que um tamanho adequado de janela seja utilizado.

Lembrando que a análise de estacionariedade é dependente de vários fatores, tamanho da janela, dimensão de reconstrução, procedimento de reconstrução do espaço de fase, e a resolução da análise (ϵ na Eq. 3.7), então todos esses parâmetros precisam ser variados para encontrar o ponto ótimo da análise.

Manuca e Savit (1996) propõem um algoritmo de agrupamento das janelas. Tal algoritmo é baseado na medida M_{ij} (Eq. 3.10) que normaliza o intervalo da dinâmica de toda a série, sendo n o número de janelas, d_{ik} a proximidade dinâmica anteriormente definida entre as janelas i e k , θ é a função *Heaviside* já mencionada anteriormente. Na proposta dos autores, $\sqrt{2}$ é o limiar usado (limiar dinâmico), mas em experimentos preliminares, verificou-se que este valor não era adequado, então este é mais um parâmetro que teve seu valor ajustado. M_{ij} próximo de zero significa que as janelas i e j têm aproximadamente a mesma distância entre todas as outras janelas, ou seja, uma possível distância entre a dinâmica efetiva destas janelas é altamente limitada.

Os seguintes passos definem o algoritmo de agrupamento utilizado:

1. Defina o número de grupos como zero ($m = 0$), coloque todas as janelas no conjunto de janelas não agrupadas ($n_u = n$, n número total de janelas), e calcule todos os M_{ij} (todos os pares de janelas);
2. No conjunto de janelas não agrupadas, calcule o número de vizinhos próximos (Eq. 3.11) de cada janela;

$$N_i = \sum_{j=1}^{N_u} \theta(\sqrt{n} - M_{ij}) \quad (3.11)$$

3. Crie um novo grupo (aumenta m em um). Pegue do grupo de janelas não agrupadas a que tem o maior número de vizinhos próximos (maior N_i), coloque-a no grupo atual G_m e a remova do grupo de janelas não agrupadas;
4. Pegue do grupo de janelas não agrupadas, aquela (W_j) com a menor média M_{ij} em relação as janelas (W_j) pertencentes ao grupo atual G_m , e com média menor que $\sqrt{n}$ (Eq. 3.12) e a coloque no grupo G_m .

$$\langle M_{ij} \rangle < \sqrt{n}, W_j \in G_m \quad (3.12)$$

Repita esse passo até que não haja nenhuma janela não agrupada que satisfaça a condição de proximidade (Eq. 3.12);

5. Repita os passos 2-4 até que todas as janelas estejam agrupadas.

Tabela 1 – Tabela de contingência

$X \setminus Y$	Y_N	Y_S	SUM
X_1	$n_{11} = 3$	$n_{12} = 1$	$a_1 = 4$
X_2	$n_{21} = 2$	$n_{22} = 1$	$a_2 = 3$
X_3	$n_{31} = 0$	$n_{32} = 3$	$a_3 = 3$
SUM	$b_1 = 5$	$b_2 = 5$	$n = 10$

Este processo resulta em m grupos (a variável m não tem relação com a dimensão de imersão do espaço de fase), é assumido que todas as janelas de um dado grupo têm a mesma dinâmica e o mesmo fenômeno que muda os parâmetros do sistema.

Analizando o resultado do agrupamento juntamente com os rótulos das janelas, é possível racionalizar sobre a estacionariedade da série temporal. Devido à disponibilidade dos rótulos, é possível avaliar a qualidade do agrupamento com o índice *Rand* ajustado (ARI) (HUBERT; ARABIE, 1985). O ARI foi usado por ser bem estabelecido na comunidade de agrupamento, e adequado quando os rótulos estão disponíveis. O ARI (Eq. 3.13) é a versão do índice *Rand* corrigido para o acaso, e seu intervalo de valores vai de valores negativos, se o índice é menor que o esperado, a +1 no caso de casamento perfeito.

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} / \binom{n}{2}}{\frac{1}{2} [\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - \sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} / \binom{n}{2}}. \quad (3.13)$$

As quantidades na fórmula são obtidas da tabela de contingência de dois processos de agrupamento. A tabela 1 fornece um exemplo de tal tabela, construída a partir da figura 8. Na figura, *label* representa o rótulo de cada janela como o resultado de um processo de agrupamento, e *clustering*, o resultado de outro processo de agrupamento, tal como o usado aqui.

No exemplo, o valor do ARI calculado é 0,15 sinalizando um agrupamento ruim, mas da perspectiva da estacionariedade, este é um bom agrupamento. Este resultado ruim acontece porque o ARI penaliza processos que geram mais grupos do que o necessário, no exemplo dois. Por esta razão, outra medida também foi utilizada na avaliação do agrupamento, a acurácia de grupo.

A acurácia de grupo é uma medida que foi definida pelo autor para compensar pela limitação do ARI.

Label	N	N	N	N	N	S	S	S	S	S
Clustering	1	2	1	1	2	3	2	3	1	3

Figura 8 – Exemplo de agrupamento

A acurácia de grupo considera um grupo atribuído erroneamente, se o grupo é considerado como pertencente a um rótulo, mas esse grupo é atribuído a outro rótulo. Um grupo é considerado pertencer ao rótulo mais frequentemente atribuído. A figura 8 ilustra isto, os grupos 1 e 2 são

considerados como pertencentes ao rótulo N , pois 1 e 2 são mais frequentemente atribuídos ao rótulo N , e o grupo 3 como pertencente ao rótulo S . É por isso que há dois marcadores em vermelho no exemplo (sétima e nona colunas) mostrando que eles foram erroneamente atribuídos. No exemplo dado, a acurácia de grupo calculada é 0,8 representando um bom agrupamento.

Abaixo são listados os parâmetros e seus valores experimentados:

- **Atraso (τ):** O atraso temporal utilizado na reconstrução do espaço de fase. O trabalho em que essa análise determinística é proposta (MANUCA; SAVIT, 1996) usa o valor fixo de um. Mas os autores afirmam que outros métodos de reconstrução podem ser utilizados, então o valor de atraso utilizado é o valor do primeiro mínimo da informação mútua calculado para cada janela (FRASER; SWINNEY, 1986);
- **Dimensão de reconstrução:** Essa análise utiliza um processo iterativo de reconstrução do espaço de fase, e os valores de dimensão utilizados vão de 1 a 15;
- **Limiar dinâmico:** O limiar dinâmico (Eq. 3.10) é o parâmetro da sensibilidade de criação de grupos, determinando se duas janelas estão próximas o suficientes para serem colocadas no mesmo grupo. O valor utilizado no trabalho original (MANUCA; SAVIT, 1996) é apenas $\sqrt{2}$, mas como os resultados com esse valor não foram satisfatórios os seguintes valores foram experimentados, 1; 0,5; 0,25; 0,1; 0,01;
- **Tamanho de janela:** Os mesmos valores do tamanho da janela utilizados na análise estocástica foram utilizados aqui, 128, 256, 512, 1.024, 2.048, e 4.096. Os mesmos valores foram utilizados para poder comparar as análises;
- **Epsilon (ϵ):** O parâmetro ϵ , utilizado na correlação integral de Grassberger-Procaccia (Eq. 3.7), teve os seguintes valores 2; 1; 0,5; 0,2; 0,1 vezes o desvio padrão em cada janela. Como as séries temporais têm escalas e amplitudes diferentes, a parametrização em termos do desvio padrão é uma boa alternativa para esse parâmetro. Esses valores foram escolhidos pois fixou-se o teto em 2, já que dois desvios padrão englobam quase todos os pontos ao redor da média. A partir do teto, foi-se diminuindo os valores de forma arbitrária para cobrir várias possibilidades.

3.2.1.3 Avaliação

A avaliação da análise de estacionariedade é feita para as 2700 execuções determinísticas (todas as combinações de parâmetros, 15 dimensões, 6 limiares dinâmicos, 6 tamanhos de janelas, e 5 *epsilons*), e 30 estocásticas (6 diferentes tamanhos de janelas e 5 níveis de significância). Como essa é uma abordagem nova, a avaliação consiste mais de uma análise exploratória. Então a análise será mais qualitativa, comparando as métricas descritas, ARI e acurácia de grupo para o caso determinístico, além do *effect size* e poder dos testes para o caso estocástico.

3.2.2 Classificação de crises epiléticas

O método geral de trabalho consiste na utilização de uma janela deslizante sobre o sinal utilizado. Para cada posicionamento das janelas, considerando uma janela para cada canal do sinal, será utilizada uma técnica de conversão de série temporal em rede complexa. E para cada rede complexa gerada são calculadas as várias métricas descritas na revisão bibliográfica. Essas métricas estão indexadas pelo número da janela e sua classe, e dão origem a um novo conjunto de dados. Esse conjunto de dados servirá de entrada para vários algoritmos de classificação. Também é realizada a seleção de atributos nesse novo conjunto de dados a fim de se selecionar os mais representativos, já que o cálculo de todos será computacionalmente caro e muitas vezes redundante.

De modo geral esse processo pode ser visualizado na figura 9.

Figura 9 – Esquema geral de predição de epilepsia. Adaptado de (MORMANN, 2008).

São utilizadas as três classes de técnicas de conversão de séries temporais em redes complexas referenciadas na revisão bibliográfica, redes de proximidade, grafos de visibilidade e redes de transição. Dentre as redes de proximidades, há três tipos de redes, redes cíclicas, redes de correlação, e redes de recorrência; no trabalho a rede de correlação não foi utilizada pois ao

final das execuções constatou-se um erro na implementação, e como os resultados desse tipo de rede estavam errados, eles não foram incluídos no texto pois não houve tempo hábil para reparar o erro e reexecutar.

E as métricas de caracterização de redes utilizadas são grau, medidas de centralidade (intermediação, autovetor, *PageRank*, proximidade e *K-core*), entropia, coeficiente de aglomeração, complexidade, dominância do ponto central e assortatividade.

Para a classificação são usadas árvores de decisão (implementação disponível no pacote *scikit-learn*²) possibilitando alguma representação simbólica dos resultados (pode ser útil para a análise de um especialista nesse domínio) e SVM com *kernel* radial pois esta é reconhecida como uma técnica que produz bons classificadores. Além dos classificadores, k-vizinhos mais próximos e LDA.

O tamanho da janela é tipicamente variado entre 10 e 40 segundos (MORMANN; ELGER; LEHNERTZ, 2006), mas Shoeb (2009) utiliza uma janela de 6 segundos com uma sobreposição entre janelas consecutivas de 5 segundos. Como o tamanho da janela ainda não é uma questão fechada neste contexto, optou-se por janelas sem sobreposição, e com tamanhos calculados dividindo-se o tamanho da série em dois, sucessivamente. Por exemplo, séries com tamanho 4.096 são analisadas com os seguintes tamanhos de janela, 2.048, 1.024, 512, e 256. Além da divisão sucessiva do tamanho da série em dois, também foram usados alguns valores da divisão em três partes.

A figura 1 ilustra o fluxo dos dados de forma mais detalhada. Conforme explicado na seção 1.3.2, a classificação segue as seguintes etapas. Primeiramente é feita a leitura da série temporal, então cada janela da série é convertida em grafo pelos vários métodos de conversão. Com o grafo construído, as métricas descritas são extraídas, e um conjunto de dados na forma atributo-valor é criado. Este conjunto é pré-processado, e então os algoritmos de classificação são utilizados gerando as matrizes de confusão. Calcula-se então as métricas de desempenho e a avaliação final é realizada.

Cada uma dessas etapas é descrita a seguir.

3.2.2.1 Leitura da série temporal

A leitura da série é uma etapa simples, depende apenas do formato do arquivo no qual os dados estão. A função de leitura das séries retorna além dos valores da série, os rótulos associados e algumas informações referentes a cada série e janela da série analisada, que são usadas na reconstrução do espaço de fase. Essas medidas são o atraso utilizado na reconstrução do espaço de fase, e o desvio-padrão, ambas as medidas para cada janela da série temporal, elas já foram pré-calculadas para vários tamanhos de janelas. No processo de leitura, é necessário informar os tamanhos de janelas que serão utilizadas na análise.

² Scikit-learn: Machine Learning in Python. Disponível em: <<http://scikit-learn.org/>>

Os tamanhos de janela utilizados no processamento dos dados da Universidade de Bonn foram 256, 303, 512, 910, 1.024, 2.048 e 2.730 pontos. Como cada série deste conjunto de dados tem 4.097 pontos, uma janela de 256 pontos produz 16 janelas, uma janela de 303 pontos produz 13 janelas, uma janela de 512 pontos produz 8 janelas, uma janela de 910 pontos produz 4 janelas, uma janela de 1.024 pontos produz 4 janelas, uma janela de 2.048 pontos produz 2 janelas, e uma janela de 2.730 pontos produz 1 janela. Totalizando, para cada série temporal analisada, 48 janelas para serem processadas.

Esses valores de janela seguem o mesmo princípio da análise de estacionariedade, mas além dos valores lá utilizados, também são utilizados aqui 303, 910, e 2.730. A escolha desses valores foi arbitrária e visa testar janelas com um sobreposição em relação aos valores pares.

3.2.2.2 Conversão da série em grafo

Como descrito anteriormente, a série é processada em janelas, e cada janela é convertida num grafo. Logo, cada série dará origem a várias instâncias do novo conjunto de dados criado, uma instância para cada janela convertida vezes o número de configurações de cada algoritmo de conversão, pois a maioria dos algoritmos tem parâmetros que controlam sua conversão.

Dos métodos de conversão descritos na seção de revisão dos métodos de conversão de série temporal em redes complexas (2.3.4), os que são utilizados neste trabalho são apresentados abaixo com seus respectivos parâmetros.

- **Redes cíclicas:** o ciclo é definido entre dois picos sucessivos, e possui apenas um parâmetro, o limiar da correlação a partir do qual dois ciclos são conectados por uma aresta. Os valores utilizados para esse parâmetro foram 0,01; 0,05; 0,1; 0,2; 0,3; 0,5; 0,7; e 0,9;
- **Grafos de visibilidade:** este método de conversão não tem parâmetros;
- **Redes de transição:** dois são os parâmetros desta abordagem, o número de intervalos utilizados para discretizar a série, e o limiar de probabilidade para conectar dois intervalos. Os valores do número de intervalos utilizados foram 100, 500, 1.000 e 2.000, e os valores do limiar de probabilidade foram 0,01; 0,05; 0,1; 0,2; e 0,3. Como esses parâmetros são independentes, a fim de cobrir todas as configurações, foram executadas 24 (4 intervalos e 6 limiares de probabilidade) conversões para cada janela de cada série analisada;
- **Redes de ϵ vizinhança:** este método tem apenas um parâmetro, o limiar ϵ que define a vizinhança de um ponto. Os valores utilizados foram 0,05; 0,1; 0,2; 0,5; 0,7; 1,0; 1,2; 1,5; 1,7; e 2,0;
- **Redes de k vizinhos mais próximos:** apenas o parâmetro k é necessário nessa abordagem e os valores utilizados foram de 1 a 16;

- **Redes de k vizinhos mais próximos adaptativos:** o parâmetro $E0$ desse método teve os mesmos valores que o parâmetro k do método anterior.

As redes de recorrência, *epsilon* vizinhança, k vizinhos mais próximos e k vizinhos mais próximos adaptativos, são construídas a partir do espaço de fase da série temporal. Isso implica que além dos parâmetros já discutidos, os parâmetros da elaboração do espaço de fase em questão estão implícitos quando estes métodos são utilizados. Os parâmetros utilizados na reconstrução do espaço de fase são os mesmos que os descritos na seção de método da análise de estacionariedade determinística (3.2.1.2). Nesta etapa, foram usados os mesmos valores dos parâmetros da reconstrução que os utilizados na análise de estacionariedade, margem de erro de 2% para o mínimo da informação mútua para o cálculo do atraso, τ , com 10 intervalos (do inglês *bins*), e 20%, no critério dos falsos vizinhos para se determinar o número de dimensões do espaço de fase reconstruído.

O número de dimensões do espaço de fase reconstruído é calculado pelo critério dos falsos vizinhos. Um falso vizinho é um ponto que para uma dada dimensão de reconstrução n , dois pontos são vizinhos, mas se a reconstrução for feita com $n + 1$ dimensões esses mesmos pontos não são mais considerados vizinhos. O valor desse parâmetro é definido como a porcentagem de pontos que são falsos vizinhos. Escolheu-se o valor de 20%, pois como o sinal é ruidoso, se fosse escolhido uma porcentagem menor, o espaço reconstruído teria um número excessivo de dimensões não refletindo alguns resultados da área que esses sistemas, neste caso epilepsia, não tem um número grande de dimensões (SAERMARK *et al.*, 1989; BABLOYANTZ; SALAZAR; NICOLIS, 1985).

Em todas as conversões de série temporal em grafo, pode acontecer da conversão gerar grafos disjuntos, e direcionados. Para simplificar a análise, todos os grafos após a conversão foram transformados em não direcionados, sem multi-arestas e *loops*, e o maior componente do grafo foi extraído. Estas operações estão disponíveis na biblioteca de manipulação de grafos *igraph* (CSARDI NAM LEE, 2014).

Com isso, cada janela de cada série temporal analisada é convertida 75 vezes (8 redes cíclica, 1 grafo de visibilidade, 24 redes de transição, 10 redes de *epsilon* vizinhança, 16 redes de k vizinhos mais próximos, 16 redes de k vizinhos mais próximos adaptativos).

3.2.2.3 Extração das métricas

Tendo o grafo que representa cada janela, é necessário extrair medidas que caracterizem esses grafos e possa então ser criado um novo conjunto de dados no formato atributo-valor. A seção 2.3.2 apresenta várias métricas utilizadas para caracterizar as redes complexas, utiliza-se 12 das métricas descritas, apenas a eficiência foi descartada pois a complexidade de tempo da implementação utilizada estava muito alta. Algumas dessas métricas são definidas para cada vértice do grafo, portanto, não é um valor único que caracteriza o grafo. Nesses casos,

como medidas, são extraídas a média e a variância. Este foi o caso para as seguintes métricas: grau dos vértices, coeficiente de aglomeração, grau médio dos vizinhos (k_{nn}), centralidade por intermediação, centralidade por autovetor, centralidade por proximidade, centralidade por *k-core* (*coreness*), *PageRank*. Isso totaliza 21 medidas que caracterizam cada janela, são elas: complexidade, transitividade, dominância do ponto central, média e variância do *PageRank*, média e variância da intermediação, média e variância do *coreness*, média e variância da centralidade por autovetor, média e variância do grau médio dos vizinhos, entropia do grau, média e variância do coeficiente de aglomeração, média e variância da proximidade, média e variância do grau e assortatividade. Essas 21 medidas são os atributos do novo conjunto de dados.

Dessas medidas, apenas a dominância do ponto central e a entropia do grau não estão disponíveis na biblioteca *igraph*, mas com uma simples manipulação algébrica de outras funções disponíveis na biblioteca é possível calculá-las.

O arquivo final com o conjunto de dados é salvo no formato *comma-separated value* (CSV).

Cada conjunto de dados gerado é definido como a conversão de todas as séries temporais analisadas, fixado um método de conversão e uma configuração de parâmetros, e fixado também um tamanho de janela para analisar essas séries. Isso totaliza 75 conversões considerando os algoritmos de conversão e seus parâmetros, vezes os 7 tamanhos diferentes de janelas, ou seja, são 525 conjuntos de dados diferentes.

3.2.2.4 Pré-processamento

No contexto deste trabalho o pré-processamento é considerado a etapa após a conversão da série em grafo e antes da classificação.

Para cada um dos 525 conjuntos de dados gerados, é feita a divisão de 20% do total reservado para teste e 80% reservado para o treinamento, essa divisão é feita através de uma amostragem aleatória.

Como várias medidas têm algum tipo de redundância, elas acabam medindo características sobrepostas. A fim de diminuir a dimensionalidade e reduzir essa redundância, é feita a seleção de características. Como o foco do trabalho não está no estudo de técnicas de seleção de características, optou-se pela escolha da técnica classificação de características pela eliminação recursiva de características e validação cruzada do número de melhores características (RFECV), disponível na biblioteca *Scikit-learn* (PEDREGOSA *et al.*, 2011).

Dado um estimador externo que atribui pesos às características, a eliminação de características recursiva (RFE) seleciona características considerando recursivamente conjuntos cada vez menores das características. Primeiro, o estimador é treinado com o conjunto inicial de características, e pesos são atribuídos a cada uma delas. Então, as características com os menores pesos são removidas do conjunto de características. Este procedimento é repetido recursivamente

até que o número desejado de características seja alcançado. RFECV realiza o RFE dentro da validação cruzada de modo a encontrar o número ótimo de características para essa abordagem.

RFE não garante uma solução ótima, é um método guloso sub-ótimo, mas há indícios de que este método é mais robusto ao sobre-treino (*overfitting*) que outros métodos, incluindo a busca combinatória (GUYON *et al.*, 2002).

A implementação disponível tem os seguintes parâmetros: estimador, passo, parâmetros da validação cruzada e métrica de avaliação da classificação. O estimador é o algoritmo de classificação que atribui pesos às características. Não são todos os algoritmos de classificação que podem ser utilizados, apenas os que retornam pesos às características, por exemplo, o algoritmo Knn não pode ser utilizado.

Neste trabalho utilizou-se um comitê de *ExtraTreesClassifier* (PEDREGOSA *et al.*, 2011; GEURTS; ERNST; WEHENKEL, 2006). *Extra-trees* diferem das árvores de decisão clássicas na maneira que são divididas. Para selecionar a melhor divisão, divisões aleatórias são feitas para cada um das *max_features* aleatoriamente selecionadas, e a melhor divisão dentre elas é a escolhida. O parâmetro *max_feature* é o número de características considerado quando a divisão aleatória for feita. Neste trabalho o valor escolhido foi o padrão da implementação, a raiz quadrada do número total de características ($\sqrt{n_features}$). Outro parâmetro necessário é a escolha da função que mede a qualidade da divisão, foi utilizado a medida *Gini* de impureza. Os outros parâmetros foram deixados com seus valores padrão.

Devido à natureza aleatória esse comitê é composto por 1.000 instâncias desse classificador.

Uma validação cruzada com 10 *folds* foi aplicada ao estimador acima descrito no conjunto de dados de treinamento. Foram utilizadas duas métricas para guiar a seleção das características, a acurácia e a área sobre a curva ROC (AUC). Escolheu-se a AUC pois ela pode ser entendida como um balanço entre a sensibilidade e a especificidade, que são as métricas de avaliação mais importantes na área do trabalho.

O último parâmetro a ser utilizado pela RFECV é o passo, que estabelece o número de características removidas a cada iteração; foi utilizado um passo de 1. Embora o custo computacional seja maior com esse valor de passo, mais subconjuntos de atributos são testados quanto ao seu poder de discriminação na classificação, resultando num melhor subconjunto de atributos.

3.2.2.5 Classificação

Após o pré-processamento ocorre a etapa de treinamento, em que os classificadores têm seus modelos induzidos a partir dos conjuntos de treinamento, e estimativas de desempenho dos mesmos são calculadas.

Todos esses classificadores estão implementados na biblioteca *Scikit-learn* (*LinearDiscri-*

minantAnalysis, *DecisionTreeClassifier*, *SVC*, *KNeighborsClassifier*), e os parâmetros descritos são os necessários para a execução dos códigos da biblioteca. A fim de cobrir bem o espaço de hipóteses, os seguintes classificadores com vieses diferentes foram utilizados:

- **LDA:** Classificador linear que busca uma transformação dos dados de forma a melhor separar as classes mantendo as amostras de uma mesma classe as mais próximas possível. O parâmetro *solver* usou a decomposição por valores singulares (*svd*). Para o conjunto de dados das redes de transição esse parâmetro apresentou problemas, sendo utilizado nesse caso a solução de mínimos quadrados (*lsqr*). Os demais parâmetros do método foram deixados com seus valores padrão;
- **Árvore de decisão:** Classificador que realiza cortes, ou separações, ortogonais no espaço dos atributos. Com exceção do parâmetro *class_weight*, cujo valor foi *balanced* para ajustar automaticamente o pesos das classes, os parâmetros deste método usaram seus valores padrão;
- **SVM com *kernel RBF*:** Classificador não linear reconhecido por apresentar bom desempenho. Os parâmetros deste classificador foram *kernel* com o valor *rbf*, *gamma* com valor *auto*, que segundo a documentação do *Scikit-learn* significa usar o valor um sobre o número de atributos ($1/|atributos|$), *cache_size* com valor 10.000 para melhorar o desempenho com um conjunto de dados grande, *decision_function_shape* com valor *ovr*. Para o parâmetro *C*, que ajusta a penalidade do termo de erro, os seguintes valores foram usados 0,2; 0,5; 0,7; 1; e 1,2. Para os outros parâmetros não citados os valores padrão foram usados;
- **Knn:** Classifica novos dados baseado na distância entre os dados usados no treinamento e o dado de consulta. Os parâmetros *metric* e *p* foram configurados com os valores *minkowski* e 2, respectivamente (o que significa que a distância euclidiana é usada), já o parâmetro *algorithm* teve como valor *auto*. Para variar o número de vizinhos, o parâmetro *n_neighbors* foi variado de 1 a 15, enquanto que o parâmetro *weights*, que especifica o peso da distância dos vizinhos, teve como valores *uniform* e *distance*. O valor *uniform* trata que todos os vizinhos têm mesmo peso na determinação do rótulo do ponto sendo consultado, já o valor *distance*, pondera a importância dos vizinhos pelo inverso da distância em relação ao ponto sendo consultado.

Considerando os classificadores e seus parâmetros no total são induzidos 37 modelos para cada conjunto de treinamento.

A ideia foi tentar obter classificadores evitando o favorecimento de algum deles devido à possibilidade de determinada conversão de série em grafo, favorecer algum algoritmo devido às características intrínsecas como o seu viés, mas sem fazer uma busca exaustiva do melhor

classificador para cada conjunto de dados. Sem isso, ou seja, sem a escolha de vários classificadores e sem variar seus parâmetros, uma comparação das medidas de desempenho dos mesmos com as relatadas na literatura ficaria enviesada. Pois em geral os trabalhos divulgados têm como objetivo melhorar o desempenho em relação a abordagens anteriores, logo, eles procuram fazer vários ajustes nos parâmetros de seus métodos. Como o objetivo do presente trabalho não é obter o melhor classificador ou melhorar abordagens anteriores, não é necessário uma busca detalhada pelo melhor classificador. Busca-se apenas ter um bom classificador com as abordagens de conversão de série temporal em grafo.

O objetivo desta etapa é ter como resultado o classificador, dentre os descritos acima, com a melhor medida de desempenho. As medidas de desempenho utilizadas foram a acurácia e AUC. Para cada um dos conjuntos de dados de treinamento gerados e pré-processados, foram calculadas as estimativas da acurácia e AUC de cada um dos classificadores utilizando uma validação cruzada de 10 *folds*. Resultando em 37 modelos de classificadores, baseado na média da métrica de desempenho, acurácia e AUC, da validação cruzada foi escolhido o melhor modelo para cada conjunto de dados.

A escolha dessa validação cruzada é baseada no trabalho de Kohavi (1995), no qual verifica-se que a seleção do modelo feita com 10 *folds*, tendo a acurácia como estimativa de desempenho, é melhor do que o *leave one out*.

Como o conjunto de dados da Universidade de Bonn pode ser visto como possuindo 5 classes, escolheu-se fazer a seguinte separação. Criou-se 5 conjuntos de dados diferentes a partir do original, para 5 cenários possíveis. Selecionando apenas as séries dos conjuntos O (pacientes saudáveis), N (pacientes com epilepsia no período interictal no hemisfério oposto ao do foco da crise) e F (pacientes com epilepsia no período interictal na região do foco da crise), tenta-se avaliar a classificação de sinais de pacientes saudáveis e pacientes com epilepsia mas fora da crise. Selecionando as séries dos conjuntos O e S (pacientes com epilepsia durante a crise), avalia-se a distinção de um sinal normal e um sinal durante a crise. Com as séries dos conjuntos N e S, tenta-se avaliar a distinção entre um sinal interictal mas fora da região focal e um sinal durante a crise. O mesmo para as séries F e S, mas agora comparando um sinal da região focal com a crise. E por fim, contrapondo os dois conjuntos de séries do período interictal, conjuntos N e F, com as séries durante a crise.

Além de todas as variações anteriormente mencionadas, há essa diferenciação entre esses 5 cenários de classes diferentes. Embora alguns desses cenário tenham séries de três conjuntos diferentes, a classificação será sempre binária, dependendo da combinação a classe positiva será composta pelas instâncias do conjunto O (pacientes saudáveis), ou do conjunto S (durante a crise). No caso em que os conjuntos O e S estejam juntos, a classe positiva será composta pelo conjunto S.

3.2.2.6 Avaliação

Ao final do procedimento descrito até aqui, são produzidos um classificador para cada conjunto de dados gerado, 525 no total, vezes o número de funções objetivo dos classificadores, acurácia e AUC, vezes o número de cenários de classes diferentes, que são 5. Com isso, são gerados 5.250 modelos, cada um desses modelos é o de melhor desempenho dentre os descritos na seção anterior (3.2.2.5) no conjunto de treinamento. Como descrito na seção que revisa os métodos de classificação de epilepsia da literatura (2.1.3), a sensibilidade e especificidade são as métricas objetivo no contexto do trabalho, e como a curva ROC faz um balanço entre essas medidas a AUC é a medida natural para considerar o desempenho. A acurácia também foi incluída pois vários trabalhos a utilizam como medida de desempenho, portanto, a fim de comparar com a literatura ela é usada.

Agora é descrito o processo de avaliação geral desses modelos em relação aos objetivos do trabalho.

O primeiro passo será avaliar dentre cada algoritmo de conversão, 6 tipos, quais as melhores configurações de cada algoritmo. Na fase de treinamento, por meio da validação cruzada, é calculada uma estimativa de desempenho, essa medida é usada para a seleção do melhor modelo. Nesta etapa, de avaliação, o desempenho é calculado no conjunto de dados separado para o teste dos modelos.

Como os dados de teste não são usados na etapa de treinamento, a estimativa de desempenho deles é não enviesada, mas é uma estimativa pontual o que dificulta a comparação entre as medidas de desempenho. Portanto, é executado o processo de *bootstrap* a fim de se obter um intervalo de confiança para essas medidas de desempenho.

Depois de selecionados os melhores modelos de cada algoritmo de conversão, outras métricas também serão analisadas: especificidade, sensibilidade e a curva ROC.

Dada uma matriz de confusão de um classificador, como a representada na figura 1, é possível calcular as métricas necessárias à análise. A matriz é composta pelas entradas verdadeiro positivo (TP), falso negativo (FN), falso positivo (FP), verdadeiro negativo (TN). Abaixo são descritas as equações para a acurácia, sensibilidade e especificidade.

- Sensibilidade = $\frac{TP}{TP+FN}$ (mede o quão boa é a detecção dos positivos);
- Especificidade = $\frac{TN}{TN+FP}$ (mede o quão bom o classificador é em evitar falso alarmes);
- F1 score = $\frac{2*TP}{2*TP+FP+FN}$ - (Média harmônica entre *precision* e *recall*);
- Acurácia = $\frac{TP+TN}{TP+FP+TN+FN}$.

A curva *Receiver operating characteristics* (ROC) é um gráfico que ilustra o desempenho do classificador conforme um limiar de classificação é variado. A curva é grafada pelos valores

das taxas de verdadeiro positivo pela taxa de falso positivo conforme esse limiar é alterado. A área sobre a curva ROC (AUC) é uma medida de desempenho que avalia de maneira geral o classificador. [Fawcett \(2006\)](#) traz uma introdução a análise de curvas ROC.

[Mormann, Elger e Lehnertz \(2006\)](#) destacam ainda uma métrica interessante de avaliação, a especificidade interictal. Essa métrica avalia a porção do tempo interictal que os verdadeiros negativos são detectados. Em outras palavras, porção do tempo interictal em que o paciente não está esperando um falso alarme.

Há a crítica que nesse domínio, muitos trabalhos não comparam suas abordagens com um preditor aleatório, mas também não definem o que é um preditor aleatório. Um preditor aleatório neste trabalho, é definido como um preditor cujos rótulos da classe positiva são atribuídos aleatoriamente com a probabilidade igual à porcentagem de instâncias da classe positiva, e o mesmo para a classe negativa. Esse preditor aleatório tem seu desempenho comparado com os classificadores selecionados.

Os modelos selecionados são então comparados com os resultados relatados na literatura.

RESULTADOS

Esta seção descreve e discute os resultados obtidos seguindo os métodos da seção 3.2. Primeiramente serão apresentados os resultados para a análise de estacionariedade, e em seguida os resultados da classificação de sinais de epilepsia.

4.1 Análise de estacionariedade

As figuras 10, 11, e 12 resumizam os resultados da execução nos três conjuntos de dados, sinal *gaussiano*, sinal de *Lorenz*, e EEG. Cada uma contém a média de uma das métricas de avaliação, ARI ou acurácia de agrupamento, calculadas sobre todos os outros parâmetros. Por exemplo, nas imagens com o eixo x tamanho da janela (*window size*), para um dado valor do tamanho da janela, o ponto no gráfico é a média calculada sobre todas as combinações dos outros parâmetros (todos os limiares dinâmicos, dimensões e *epsilons*). O mesmo raciocínio é aplicado aos outros tipos de gráficos.

A primeira linha das figuras 10 e 11 apresentam os resultados para o índice ARI, e a linha de baixo apresenta os resultados da acurácia de agrupamento. A figura 12 também está dividida em duas linhas, a de cima mostra as medidas de desempenho, e a de baixo, a média do poder do teste de hipótese para cada nível de significância.

Como a média é tomada com relação à variação de todos os outros parâmetros, o valor dessa média incorpora possíveis resultados ruins devido a valores inadequados dos outros parâmetros. Mas como estes possíveis valores inadequados são utilizados em todos os valores do parâmetro cuja média está sendo calculada, é possível comparar essas médias em termos relativos. Ou seja, é possível inferir qual os melhores valores para o parâmetro em questão.

Em todas as execuções determinísticas, o índice ARI tem valores pequenos. Este foi o motivo que levou ao uso de outra métrica, lembrando que o valor máximo do ARI é +1. Esses valores pequenos ocorrem pelo provável excessivo número de grupos gerados. E como mostrado

Figura 10 – Resultados da análise determinística (window size and dynamic threshold).

no final da seção 3.2.1.2 quando o ARI foi apresentado, ele tem seu valor reduzido se o número de grupos gerados pelo processo de agrupamento for muito diferente do número de rótulos.

No caso determinístico, a tendência nos gráficos do tamanho da janela parece ter alguma semelhança, embora a curva de *Lorenz* seja o menos similar. Em geral, o desempenho para 128 pontos não é tão bom quanto poderia ser, isso pode ser explicado pois esse é um tamanho de janela pequeno. Como a reconstrução do espaço de fase é feita em cada janela, e quando uma janela deste tamanho tem seu espaço de fase reconstruído, isto resulta em poucos pontos, ou seja, um espaço muito esparso. Logo, o cômputo das medidas usadas no processo de agrupamento apresenta um desempenho ruim. O pico de desempenho parece ocorrer ao redor de 512 pontos para o EEG e o sinal gaussiano, e ao redor de 1.024 ou 2.048 pontos para o sinal de *Lorenz*. A janela de 4.096 pontos tem o melhor resultado para *Lorenz* segundo o ARI, provavelmente devido

Figura 11 – Resultados da análise determinística (epsilon and embedding dimension).

ao número de grupos retornados. Com 4.096 pontos são gerados no máximo, dois grupos. A variação nos resultados do sistema de *Lorenz* torna explícita a importância do parâmetro tamanho da janela em sistemas determinísticos.

Quanto ao limiar dinâmico, claramente o primeiro valor (0,001) considerando a acurácia de grupo e o segundo valor (0,1) considerando o ARI têm os melhores desempenhos. Este é um parâmetro exclusivo do processo de agrupamento, responsável pela sensibilidade do método na geração de novos grupos. Como o ARI é fortemente influenciado pela diferença entre o número de grupos e o número de rótulos, isto ajuda a explicar porque o segundo, e não o primeiro menor valor, obteve o melhor resultado.

A tendência crescente é notável nos gráficos do parâmetro *epsilon*. Este parâmetro impacta no cálculo da correlação integral, que fornece uma noção da probabilidade de encontrar

Figura 12 – Resultados da análise estocástica.

pontos próximos entre si no espaço de fase. Logo, para valores pequenos, esta probabilidade acaba sendo próxima de zero, e não é possível construir a estrutura subjacente do espaço de fase. Para valores grandes, o método é capaz de conectar os pontos e construir tal estrutura. Como este parâmetro não afeta diretamente o número de grupos, a tendência nos gráficos das duas medidas é similar.

A dimensão de reconstrução tem uma tendência para baixo em todos os conjuntos de dados em ambas as métricas. Interessante o fato do valor de dimensão 3, o qual é a dimensão real do sistema de *Lorenz*, ter o melhor resultado segundo o ARI, mas não segundo a acurácia de grupo. Considerando a suposição de que um espaço de fase com muitas dimensões implica num sistema mais complexo, e à luz da navalha de Occam, parece razoável que quando da análise destes tipos de sinais, não usar uma dimensão de reconstrução maior que 5.

Considerando métodos que requerem estacionariedade e usam espaço de fase ou concei-

tos similares, de acordo com os resultados aqui encontrados, é recomendado utilizar os seguintes intervalos para os parâmetros: tamanho de janela entre 512 e 1.024 pontos, *epsilon* entre 1 e 2 vezes o desvio padrão de cada janela, e dimensão de reconstrução até 5. Esta é uma recomendação para sinais similares aos utilizados aqui, bem como métodos de trabalho parecidos.

Quanto à análise estocástica, o uso de um nível de significância muito rígido implica em poucas rejeições da hipótese nula, e como no cenário testado, e mesmo em geral, há poucas mudanças de comportamento, isto é refletido no alto desempenho apresentado. Este efeito é claramente visto no conjunto de dados gaussiano pelo seu comportamento quase linear entre sua medida de desempenho e a significância, já que a influência dos vários valores do tamanho da janela é menor que nos outros dois conjuntos de dados. Com uma significância de 0,01 há mais rejeições da hipótese nula devido a flutuações estatísticas, explicando o maior índice de desempenho para esse valor de significância já que captura as 19 transições de comportamento do sinal *gaussiano*, enquanto que uma significância de 0,3 pode não ser capaz de rejeitar a hipótese nula nas mudanças de comportamento. O mesmo padrão não é tão claro nos dados do EEG e *Lorenz*, pois o efeito da janela nesses conjuntos de dados é maior, mascarando o efeito da significância.

Assim como aconteceu na análise determinística, o tamanho da janela causou uma grande diferença no desempenho do EEG e *Lorenz*. Da mesma forma que antes, este parâmetro afeta o número de dados disponível para o método, e como o histograma influencia o teste de hipótese, esta é a provável razão desta grande influência nas medidas de desempenho.

Uma significância de 0,1 parece ser uma boa escolha, pois não é uma significância muito rígida mas mantém um bom poder do teste de hipótese.

Além dos resultados agregados mostrados, foi gerada uma grande quantidade de gráficos com as medidas da proximidade dinâmica simétrica (Eq. 3.6) e da sua versão normalizada (matriz M_{ij} , Eq. 3.10) para cada combinação de parâmetros testada na abordagem determinística. Mas devido a enorme quantia de informação presente neles, eles não foram reportados no texto. O mesmo aconteceu no caso estocástico, muita informação produzida nos experimentos não está presente no texto devido ao grande volume. Para cada série temporal testada foram gerados gráficos para cada tamanho de janela, o histograma de cada janela ao longo da série com as estatísticas do teste de hipótese, além dos valores de média, desvio padrão, *effect size* e poder do teste estatístico. O texto contém o resumo dessas informações, mas o autor pretende compilar esses dados de outras maneiras e usar isso para a submissão de artigos em periódicos.

4.2 Classificação de crises epiléticas

Essa seção apresenta os resultados dos experimentos com o objetivo de classificar sinais de epilepsia.

Tabela 2 – Informações gerais das conversões

Algoritmo	Tamanho	Número de registros	Número de registros com valor inválido
aknn	102 MB	359.280	26
knn	11 MB	359.280	58
eps	69 MB	246.672	106.604
cycle	56 MB	191.616	551
trans	92 MB	502.992	1
vis	7 MB	23.520	0

Como nesse trabalho é feita a comparação do desempenho de classificação para vários algoritmos de conversão de uma série temporal em grafo, e esses algoritmos têm vários parâmetros diferentes, optou-se por em geral apenas diferenciar as execuções em relação à métrica de desempenho, acurácia e AUC, ao tamanho da janela, e às classes das instâncias. Assim, é possível manter a comparação entre os vários algoritmos nos fatores em comum.

Nas próximas seções serão apresentados os resultados que visam mostrar como os objetivos do trabalho foram atingidos. A estrutura de seções a seguir reflete as etapas descritas na seção 3.2.2 com exceção da primeira, Leitura da série temporal.

4.2.1 Conversão da série em grafo

A tabela 2 mostra o resumo de algumas informações das conversões do conjunto de séries temporais originais em grafos. A primeira coluna estabelece qual o algoritmo de conversão foi utilizado, *aknn* é referente à rede de vizinhos mais próximos adaptativos, *knn* à rede de vizinhos mais próximos, *eps* à rede de *epsilon* vizinhança, *cycle* às redes cíclicas, *trans* às redes de transição, e *vis* ao grafo de visibilidade. A segunda coluna mostra a quantidade de memória que o arquivo com os valores extraídos das redes no formato CSV ocupa. Na terceira é apresentado o número de registros que o novo conjunto de dados tem, e na quarta é exibida a quantidade de registros com algum valor inválido. Valores inválidos ocorrem devido a algum problema numérico no cálculo, como *Not-a-Number* (NaN), ou pela definição da medida, por exemplo, infinito pode ser retornado por alguma medida e é considerado um valor inválido.

A quantidade de registros que cada algoritmo tem é devida ao número de conversões diferentes que foram realizadas. O número de parâmetros e a quantidade de valores testadas para cada parâmetro determinam esses valores. Em termos da quantidade de registros, todas as conversões produziram muitos, que serão utilizados na classificação. Nota-se que o algoritmo de *epsilon* vizinhança tem muitos registros com valores inválidos, quase sua totalidade ocorrem na medida do coeficiente de aglomeração. Provavelmente isso ocorreu devido a um valor inadequado de *epsilon*.

Neste ponto, uma análise interessante de ser feita é o estudo da distribuição do número de vértices e arestas dos grafos que cada algoritmo e valor de parâmetros geraram. É interes-

sante, pois as medidas extraídas posteriormente caracterizam a organização da rede, e a melhor comparação da organização de duas redes é dada fixando-se o número de vértices e arestas de ambas. Embora neste trabalho não se focou em garantir que duas redes tenham o mesmo número de vértices e arestas, pois o interesse está no valor utilizado na conversão e não no número de vértices e arestas gerados, o estudo dessas distribuições é interessante e pode prover algum *insight* que explique o desempenho final.

Como há muitos parâmetros de conversão resultando em muitas possibilidades de análise, apenas as conversões que apresentaram os melhores resultados foram reportadas. Mais detalhes sobre os vértices e arestas são detalhados na seção 4.2.5 e no apêndice A. Esta análise do número de vértices e arestas foi realizada após a defesa desta dissertação como uma sugestão de um membro da banca avaliadora.

4.2.2 Extração das métricas

A figura 13 mostra um exemplo das 13 medidas calculadas após uma janela de uma série temporal ser convertida num grafo. Dessas, oito têm seus valores calculados em cada vértice, logo para sumarizar essas medidas para o grafo todo, seu histograma é calculado. Desse histograma são extraídos sua média e variância como os valores finais do novo conjunto de dados gerado. Os 5 valores apresentados na tabela contida na figura já têm seus valores como escalares, calculados para todo o grafo, então são diretamente colocados no novo conjunto de dados. Esta figura é apenas para mostrar como as métricas de cada grafo são calculadas, mas é impraticável tentar mostrar todas essas informações, ou analisar as medidas dessa forma, então para ter uma ideia geral da distribuição das medidas, os *boxplots* para todas as medidas e para todos os algoritmos de conversão são exibidos nas figuras 82 a 87 do apêndice B.

No gráfico mostrado abaixo, figura 13, as medidas têm seus nomes abreviados: *compl* é a complexidade, *trans* é a transitividade, *cpd* é a dominância do ponto central, *pagerank_mean* é a média do *PageRank*, *pagerank_var* é a variância do *PageRank*, *betweenness_mean* é a média do grau de intermediação, *betweenness_var* é a variância do grau de intermediação, *coreness_mean* é a média da centralidade por *K-core*, *coreness_var* é a variância da centralidade por *k-core*, *eigenvector_mean* é a média da centralidade por autovetor, *eigenvector_var* é a variância da centralidade por autovetor, *knn_mean* é a média do grau médio dos vizinhos, *knn_var* é a variância do grau médio dos vizinhos, *dg_entropy* é a entropia do grau, *clus_coef_mean* é a média do coeficiente de aglomeração, *clus_coef_var* é a variância do coeficiente de aglomeração, *closeness_mean* é a média da centralidade por proximidade, *closeness_var* é a variância da centralidade por proximidade, *degree_mean* é a média do grau, *degree_var* é a variância do grau, e *assort* é a assortatividade.

Figura 13 – Medidas de uma janela de uma série do conjunto F utilizando uma rede cíclica.

4.2.3 Pré-processamento

A partir dos novos conjuntos de dados, é feita a separação entre conjunto de treino e de teste. No conjunto de treino é realizada a seleção de atributos como descrita na seção 3.2.2.4.

É difícil tirar conclusões generalistas com relação aos atributos selecionados, cada algoritmo de conversão tem um viés diferente e isto é refletido no comportamento das métricas. Uma discussão maior sobre a seleção de atributos realizadas está presente no apêndice C.

4.2.4 Classificação

Tendo feita a seleção de atributos para cada parâmetro do algoritmo de conversão, é realizada a etapa de classificação. Essa etapa realiza uma validação cruzada com 10 *folds* em cada classificador induzido do banco de 37 classificadores. A validação cruzada visa estimar o

desempenho do classificador. Essa estimativa, média e desvio padrão, é posteriormente utilizada para a seleção do melhor classificador para um dado conjunto de parâmetros. Novamente, a grande quantidade de informação gerada força a análise aos fatores comuns, esses fatores são as métricas utilizadas como medida objetivo do classificador, acurácia ou AUC, e os tamanhos de janela. Portanto, dentre os melhores classificadores para uma dada parametrização da conversão, é selecionado o melhor para um dado tamanho de janela, isso reduz a saída final desta etapa a 14 classificadores, um para cada tamanho de janela nas duas métricas de avaliação, como ilustram as figuras 14 a 31.

Conforme a tabela 2 e os *boxplots* (figuras 82 a 87) evidenciam, existem medidas com valores inválidos. Como alguns algoritmos não conseguem lidar com esses valores, eles precisam ser removidos do conjunto de dados. Se o número de registros com valores inválidos for menor que 10% do total de registros, então os registros com valores inválidos são removidos, caso contrário, as colunas com os valores inválidos são removidas.

Apesar da remoção dos valores inválidos, houve casos de problemas numéricos durante a indução de classificadores, o LDA por vezes produzia coeficientes com valores infinito ou *NaN*, ocasionando problemas. Para esses casos foi utilizado um *DummyClassifier* apenas para evitar problemas no fluxo automático de processamento, esse classificador rotula as amostras baseado na proporção das classes. Para evitar a seleção deste classificador como um dos melhores, o valor de sua medida de desempenho na seleção dos melhores foi fixada em zero.

A seleção do melhor classificador para cada parâmetro de conversão segue o critério de: primeiro ordenar pelo maior valor de desempenho, depois por uma preferência de tipo de classificador, LDA, árvore de decisão (DT), Knn, SVM (SVC), e *Dummy*, nesta ordem, e por fim, pelo menor valor de desvio padrão. O critério de preferência do tipo do classificador visa a seleção de classificadores com uma complexidade menor primeiro; isso é um pouco artificial pois a complexidade depende dos parâmetros utilizados para induzir o modelo, mas como esse não é o primeiro fator da ordenação, esse viés é amenizado.

O processo descrito acima foi realizado para cada algoritmo de conversão e para cada configuração de classe, NS, FS, OS, ONF e NFS. Cada figura de 14 a 31, apresenta o resultado em forma de gráfico e tabela para cada uma dessas variantes. A tabela da figura mostrada também apresenta o número de atributos utilizados pelo classificador bem como alguns parâmetros específicos de cada modelo. Para a árvore de decisão é mostrado o número de nós; o Knn tem mostrados o valor de K e tipo de peso da distância, *uniform* ou *distance*, o SVM tem como parâmetros o valor dos coeficientes de regularização C, o valor do *gamma* (definido como *auto* em todas as execuções), o *kernel* (também definido como *rbf* em todas as execuções), e o número de vetores de suporte. O número de nós de uma DT, o número de vizinhos do Knn e o número de vetores de suporte da SVM podem ser vistos como medidas da complexidade do classificador. Cabe ressaltar que cada célula da tabela é um classificador diferente, então, mesmo que duas células adjacentes tenham o LDA com medidas de desempenho iguais, as células representam

dois classificadores diferentes que por acaso são do mesmo tipo e têm o mesmo desempenho.

Os trabalhos que utilizam o conjunto de dados da Universidade de Bonn, em geral utilizam apenas os sinais dos pacientes saudáveis e sinais durante a crise. As várias classes utilizadas, NS, FS, OS, ONF e NFS, visam fazer uma análise mais completa, tentando induzir classificadores em cenários que combinam sinais durante a crise e de paciente saudável, mas também entre momentos fora e durante a crise, ou mesmo, fora da crise e sinal de paciente saudável. Este último parece o mais indicado para uso clínico que tenta diagnosticar uma pessoa com epilepsia; afinal, a dúvida nesse caso é se o sinal medido da pessoa é um sinal interictal ou saudável. Outro exemplo de uso clínico seria na classe NFS, em que se tenta diferenciar o sinal de uma crise de um sinal interictal (período entre crises).

As figuras 14, 15 e 16, são referentes à etapa de treinamento das redes *aknn*. A figura 14 traz o desempenho do melhor classificador selecionado para as classes FS e NS, enquanto que a figura 15 tem resultado dos classes NFS e ONF, o desempenho está discriminado pelas métricas utilizadas, acurácia e AUC, e pelas janelas. A figura 16 traz apenas o desempenho da classe OS. O desempenho com a medida AUC foi melhor que o da acurácia, apenas em duas situações, a acurácia obteve melhor desempenho, o nível do desempenho geral da acurácia está bom. Considerando a AUC, o classificador LDA é o mais frequente em todas as classes. O fato do LDA ser o melhor classificador em geral é um bom sinal, já que é um classificador linear, o que significa que ele generaliza bem, e que suas estimativas são boas. No geral, o desempenho da AUC está alto para todas as classes; em várias situações acima de 0,9.

Ambas as métricas parecem ter uma tendência a crescer conforme o tamanho das janelas aumenta. Em geral, não parece haver muita diferença no número de atributos selecionados entre as métricas. Na acurácia, o LDA também é o classificador mais frequente, mas há uma ocorrência um pouco maior de DT. Interessante que para todas as classes, DT foi o melhor classificador para janelas de tamanho 2.730 na acurácia.

Figura 14 – Desempenho no conjunto de treino da rede *aknn* para as classes FS e NS.

Figura 15 – Desempenho no conjunto de treino da rede *aknn* para as classes NFS e ONF.

Figura 16 – Desempenho no conjunto de treino da rede *aknn* para as classes OS.

Agora são analisadas as medidas de desempenho das redes *knn*, figuras 17, 18 e 19. Assim como ocorreu com as redes *aknn*, as redes *knn* apresentam uma tendência crescente com o tamanho da janela. O desempenho em geral também foi bom, com a AUC sendo melhor que a acurácia. Novamente, o LDA foi o classificador mais frequente para a AUC, o DT o mais frequente para a acurácia.

Figura 17 – Desempenho no conjunto de treino da rede *knn* para as classes FS e NS.

Figura 18 – Desempenho no conjunto de treino da rede *knn* para as classes NFS e ONF.

Figura 19 – Desempenho no conjunto de treino da rede *knn* para as classes OS.

As figuras 20, 21 e 22 são referentes aos desempenhos das redes *eps*. A AUC apresenta um ótimo desempenho, muito próximo de 1, seguida de perto pela acurácia. O classificador mais frequente para a AUC é o LDA, exceto na classe ONF, na qual o Knn aparece mais. Vale ressaltar que em algumas figuras, os valores estão oscilando próximos de 1, mas a tabela apresenta valores fixos em 1, isso ocorre pois os valores na tabela foram arredondados para dois dígitos decimais, e a oscilação mostrada na figura é menor que essa precisão. O desempenho em geral nas redes *eps* está muito bom, muito melhor que das redes *aknn* e *knn* que têm o mesmo viés de transformação da série temporal em grafo. Talvez esse tipo de conversão seja mais adequada que os outros dois, ou que a seleção dos parâmetros aqui seja mais fina, já que o parâmetro aqui é um número real e nas outras duas redes é um número natural.

Figura 20 – Desempenho no conjunto de treino da rede *eps* para as classes FS e NS.

Figura 21 – Desempenho no conjunto de treino da rede *eps* para as classes NFS e ONF.

Figura 22 – Desempenho no conjunto de treino da rede *eps* para as classes OS.

As redes que vêm a seguir têm vieses diferentes das três anteriores, e as figuras 23, 24 e 25 trazem o desempenho para as redes *cycle*. Mais uma vez a tendência crescente em relação ao tamanho da janela está presente. O desempenho da AUC está consistentemente melhor que da acurácia, mas ambos estão muito bons, acima de 0,85. O LDA aparece empatado com o Knn como o mais frequente na classe OS, entretanto, nas outras classes apenas o LDA foi o melhor classificador selecionado. Já para a acurácia, o Knn é o mais frequente na classe OS, e sem a presença do LDA nessa classe, mas para as outras classes apenas o LDA aparece.

Figura 23 – Desempenho no conjunto de treino da rede *cycle* para as classes FS e NS.

Figura 24 – Desempenho no conjunto de treino da rede *cycle* para as classes NFS e ONF.

Figura 25 – Desempenho no conjunto de treino da rede *cycle* para as classes OS.

A seguir, as figuras 26, 27 e 28 trazem o desempenho das redes *trans*. O desempenho geral está bem inferior ao dos demais algoritmos de conversão, apenas a acurácia nas classes NFS e ONF um pouco melhor, mas ainda assim baixo em relação às outras redes. Outra diferença visível é a inversão entre acurácia e AUC como a melhor medida, só na classe FS que a AUC é melhor, mas seguida de perto pela acurácia. A AUC se mantém ao redor de 0,5, a acurácia para as classes NFS e ONF mantém um desempenho próximo de 0,67. O classificador mais frequente na AUC é o DT nas classes FS, NFS e ONF; já nas classes NS e OS, Knn é o mais frequente. No caso da acurácia, o DT também é o mais frequente nas classes FS e OS, o Knn é o mais frequente nas classes NFS e ONF, e o LDA na classe NS. As redes *trans* apresentaram o pior desempenho. Nota-se que em quase todas as execuções o número de atributos utilizado foi o máximo, o que sugere que esse tipo de rede pode produzir atributos com pouco grau de discriminação.

Figura 26 – Desempenho no conjunto de treino da rede *trans* para as classes FS e NS.

Figura 27 – Desempenho no conjunto de treino da rede *trans* para as classes NFS e ONF.

Figura 28 – Desempenho no conjunto de treino da rede *trans* para as classes OS.

O desempenho das redes *vis* mostrados nas figuras 29, 30 e 31 tem um comportamento que destoa das demais redes. As classes ONF e OS mantêm um comportamento parecido com as quatro primeiras redes, AUC melhor que a acurácia, e uma tendência crescente com o tamanho da janela. Mas parece que nas classes ONF e OS, o processo de conversão de série temporal em grafo deixa a separação dos rótulos das instâncias mais difícil, já que classificadores mais complexos foram os mais frequentes, sendo SVC o mais frequente, apenas para a acurácia na classe ONF que o DT foi o mais frequente. No geral, o desempenho de ambas as métricas foi bom, mas o comportamento nas classes FS, NS e NFS foi mais oscilatório que para os outros tipos de redes. O SVC foi o classificador mais frequente nas classes NS e NFS para a AUC, mas o LDA e DT também aparecem bastante.

Figura 29 – Desempenho no conjunto de treino da rede *vis* para as classes FS e NS.

Figura 30 – Desempenho no conjunto de treino da rede *vis* para as classes NFS e ONF.

Figura 31 – Desempenho no conjunto de treino da rede *vis* para as classes OS.

4.2.5 Avaliação

Tendo induzido os classificadores e selecionado os melhores por janela e métrica de desempenho, os resultados obtidos na etapa de classificação são avaliados. A falta de uma avaliação rigorosa é uma crítica feita na literatura, fonte de resultados conflitantes e um dos motivos de não haver uso clínico das abordagens propostas.

Como descrito na seção de métodos, a avaliação compara o desempenho dos classificadores selecionados com um preditor aleatório, em um conjunto de dados separado para teste. Essa comparação é feita utilizando 1.000 amostras de *bootstrap* do conjunto de teste. Cada amostra de *bootstrap* consiste num conjunto de dados amostrado com reposição do conjunto de teste original, portanto, são criados 1.000 conjuntos de teste. Com isto é possível ter medidas da dispersão da estimativa de desempenho, desvio padrão das métricas de desempenho, e a aplicação do teste estatístico t de *Welch* para duas amostras independentes, o classificador e o preditor aleatório. Os classificadores que passarem no teste para os níveis de significância especificados, são estatisticamente melhores que o preditor aleatório, ou seja, as estimativas de desempenho no conjunto de teste fornecem evidência que o classificador tem maior poder discriminativo que um palpite, o fato de serem dados não utilizados na etapa de treino é importante.

As amostras do conjunto de teste garantem que a estimativa obtida não seja enviesada,

isto é importante pois o conjunto de treino foi utilizado para induzir os 37 classificadores, e o melhor dentre eles foi escolhido. Essa escolha do melhor envies a estimativa, pois a mesma amostra é utilizada para treino e para a escolha do melhor, ou seja, um viés otimista é introduzido. Somente com os resultados no conjunto de teste é possível verificar se abordagens que convertem séries temporais em grafos são úteis na classificação de sinais de epilepsia.

Após esse processo de comparação com um preditor aleatório, o melhor classificador para cada algoritmo de conversão é escolhido, com esses classificadores, os algoritmos são comparados entre si, e posteriormente com as abordagens relatadas na literatura.

As figuras com as estimativas no conjunto de teste são semelhantes às da seção anterior. Mas, como já dito, as estimativas são calculadas nas amostras de *bootstrap*, e foi incluído o *effect size* do teste estatístico realizado juntamente com um marcador para saber se o classificador em questão passou no teste de hipótese e com qual nível de significância. Foram utilizados três níveis de significância, 0,1; 0,05; e 0,01, relacionados aos seguintes símbolos, *, †, e ‡, se um desses símbolos estiver grafado após o valor do *effect size*, significa que o classificador passou no teste com o respectivo nível de significância, caso não haja símbolo algum, o classificador não passou no teste de hipótese. O *effect size* pode ser visto como uma medida da distância entre as distribuições do classificador e do preditor aleatório, e portanto, sinaliza o quão melhor um modelo é em relação ao outro.

As figuras 32, 33 e 34 mostram os resultados das redes *aknn*. Em relação ao treino, a AUC apresentou desempenho geral menor, embora razoável, e todos os classificadores passaram no teste com 0,01 de significância. Já a acurácia teve um comportamento diferente do treino, por vezes seu desempenho foi melhor que da AUC. Os valores de *effect size* em todos os classificadores são considerados altos. Como o preditor aleatório utiliza a proporção das classes para fazer sua previsão e as medidas de desempenho ficaram bem acima da proporção das classes, isso explica os altos valores do *effect size*. Exceto nas classes NS e OS, que parecem exibir alguma tendência com o tamanho das janelas, as medidas de desempenho parecem atingir picos ao redor de 910 e 1.024. Esse tamanho de janela está dentro da recomendação para o tamanho da janela obtida pela análise de determinística estacionariedade, o que faz sentido já que essa abordagem utiliza a reconstrução do espaço de fase da janela.

Figura 32 – Desempenho no conjunto de teste da rede *aknn* para as classes FS e NS.

Figura 33 – Desempenho no conjunto de teste da rede *aknn* para as classes NFS e ONF.

Figura 34 – Desempenho no conjunto de teste da rede *aknn* para as classes OS.

Conforme as figuras 35, 36 e 37 mostram, o desempenho das redes *knn* assim como o das *aknn* foi diferente do treino, em várias situações a acurácia foi melhor que a AUC. Novamente todos os classificadores passaram no teste estatístico com uma significância de 0,01; e também os valores dos *effect size* foram altos. Um pico de desempenho também é distinguível ao redor de 910 e 1.024. No geral o desempenho foi alto para as duas métricas, sendo difícil dizer qual das duas se saiu melhor.

Figura 35 – Desempenho no conjunto de teste da rede *knn* para as classes FS e NS.

Figura 36 – Desempenho no conjunto de teste da rede *knn* para as classes NFS e ONF.

Figura 37 – Desempenho no conjunto de teste da rede *knn* para as classes OS.

Novamente, as redes *eps* tiveram um desempenho superior às redes *aknn* e *knn*, que possuem o mesmo viés, como pode ser visto pelas figuras 38, 39 e 40. As duas métricas tiveram desempenho próximo de 1 em todas as classes, e todos os classificadores tiveram altos *effect sizes* e passaram no teste estatístico com significância de 0,001, a única exceção foi do classificador com janela de 256 na AUC. Este foi o único classificador que teve baixo desempenho, com valor 0,5; e consequentemente não passou no teste estatístico o que também pode ser visto no valor de *effect size* negativo, o que significa que o preditor aleatório foi melhor que ele. Este resultado contrasta com o do treino, no qual este classificador teve uma AUC próxima de 1. Este é portanto, um exemplo da importância de uma avaliação dos resultados rigorosa e com dados não utilizados para o treino dos classificadores, especialmente em abordagens como a utilizada aqui, em que a mesma amostra usada para treinar o classificador foi usada para escolher o melhor entre vários.

Os classificadores mais simples não foram escolhidos porque provavelmente o SVC que é um classificador mais complexo, e com uma chance maior de sofrer de sobre-treino (*overfitting*), foi escolhido baseado no desempenho no mesmo conjunto de dados utilizado para treinar esses classificadores. O que não significa que se um classificador menos complexo fosse escolhido seu desempenho no teste teria sido tão ruim, pois se ele é menos complexo ele sofre menos de sobre-treino e consequentemente desempenha melhor em dados novos.

Figura 38 – Desempenho no conjunto de teste da rede *eps* para as classes FS e NS.

Figura 39 – Desempenho no conjunto de teste da rede *eps* para as classes NFS e ONF.

Figura 40 – Desempenho no conjunto de teste da rede *eps* para as classes OS.

Assim como no treino, uma tendência visível está presente no desempenho das redes *cycle* no conjunto de teste visto nas figuras 41, 42 e 43. Mas diferente do treino, em várias situações a acurácia foi melhor que a AUC. Aqui também todos os classificadores passaram no teste estatístico com significância de 0,01 e alto valor de *effect size*. O desempenho geral também foi muito bom, como no treino.

Figura 41 – Desempenho no conjunto de teste da rede *cycle* para as classes FS e NS.

Figura 42 – Desempenho no conjunto de teste da rede *cycle* para as classes NFS e ONF.

Figura 43 – Desempenho no conjunto de teste da rede *cycle* para as classes OS.

Segundo as figuras 44, 45 e 46, que mostram o desempenho das redes *trans*, novamente as classes NFS e ONF apresentam os melhores resultados, ainda que um pouco abaixo do treinamento. A variação de ambas as métricas está baixa, e fixa num patamar de desempenho baixo. Apesar do baixo desempenho, várias execuções passaram no teste de hipótese com significância de 0,01; provavelmente, o desvio padrão baixo e um desempenho do preditor aleatório abaixo de 0,5 são explicados porque várias execuções passam no teste com valores tão baixos de desempenho. Como verificado na etapa de classificação, quase sempre foram utilizados todos atributos, e mesmo os classificadores mais simples não conseguiram ter bom desempenho.

Talvez um ajuste melhor dos parâmetros dessa rede consiga aumentar o poder discriminativo. Como pode ser visto pelo *boxplot* da rede *trans*, figura 86, os valores de janela utilizados parecem não trazer muita informação, o que indica que os parâmetros específicos da rede podem ser mais importantes para essa rede nesse domínio.

Figura 44 – Desempenho no conjunto de teste da rede *trans* para as classes FS e NS.

Figura 45 – Desempenho no conjunto de teste da rede *trans* para as classes NFS e ONF.

Figura 46 – Desempenho no conjunto de teste da rede *trans* para as classes OS.

É possível ver nas figuras 47, 48 e 49 que o desempenho para as duas métricas das redes *vis* foi bom em todas as classes, com a classe NFS apresentando maior consistência no desempenho para a acurácia, e a classe ONF tendo maior consistência para a AUC. Nas classes, NS, NFS, e ONF a acurácia tem um desempenho melhor que a AUC. Na classe OS a AUC, exceto no pico, tem melhor desempenho que a acurácia, na classe FS há uma oscilação entre qual das métricas é melhor, mas o pico de desempenho ocorre para a acurácia.

Houve três situações em que para a AUC, o desempenho ficou em 0,5. Dessas, apenas uma passou no teste de hipótese, mas todas tiveram um *effect size* muito baixo ou negativo. Isto mostra a importância de se ter mais informações além do valor p ou do nível de significância, pois apesar de ter passado no teste, a diferença entre as duas distribuições, classificador e preditor aleatório, é mínima.

Figura 47 – Desempenho no conjunto de teste da rede *vis* para as classes FS e NS.

Figura 48 – Desempenho no conjunto de teste da rede *vis* para as classes NFS e ONF.

Figura 49 – Desempenho no conjunto de teste da rede *vis* para as classes OS.

Tendo os resultados para cada rede, e em cada configuração de classe, é possível filtrar os melhores classificadores para cada classe, métrica, e tipo de rede. Apenas os classificadores que passaram no teste de hipótese contra o preditor aleatório foram incluídos, e o critério de ordenação foi semelhante ao anteriormente utilizado; primeiro considerou-se o maior desempenho, depois o classificador menos complexo, o menor número de atributos, o menor tamanho de janela, e por fim o menor desvio padrão.

Nas tabelas 3 a 7 são listados os melhores classificadores, e as medidas de desempenho são as calculadas no conjunto de teste. A coluna *Função objetivo* é referente à métrica que o classificador otimiza, a coluna *Rede* traz o nome da rede bem como o parâmetro de conversão utilizado, a coluna *Classificador* traz informações sobre o classificador selecionado como o melhor de acordo com o critério acima descrito, a coluna *Atributos* tem o número de atributos utilizado pelo respectivo classificador, a coluna *Amostras* tem o número de amostras do conjunto de teste, a coluna *Janela* traz o tamanho da janela utilizada, as próximas 6 colunas trazem as métricas de desempenho, acurácia (ACC), área sobre a curva ROC no conjunto de teste (AUC teste), média da área sobre a curva ROC no *bootstrap* (AUC *bootstrap*), sensibilidade (SEN), especificidade (SPC), e *F1 score* (F1). As medidas da tabela têm seus valores arredondados com duas casas decimais.

Se as colunas *Classificador* e *Atributos* não tiverem informação significa que nenhum classificador foi escolhido; esse foi o caso com a rede *trans* na classe NS.

É notável o conflito de valores da AUC no conjunto de teste e no *bootstrap*. No processo de compilação desses resultados, foram comparados os valores das medidas finais no conjunto de teste e no *bootstrap*, e constatou-se que as métricas ACC, SEN, SPC e F1 têm valores muito próximos entre o conjunto de teste e o *bootstrap*; a exceção é a AUC teste. A causa dessa divergência foi o fato de no conjunto de teste, a AUC ser calculada utilizando valores de probabilidades diferentes no teste e no *bootstrap*. No teste utilizou-se uma função de probabilidade implementada pela biblioteca e diferente para cada classificador. Alguns classificadores retornam a probabilidade baseada na distância da instância consultada ao plano hiper-separador das classes, outros classificadores realizam uma validação cruzada para estimar a probabilidade da instância. Já no *bootstrap* a probabilidade foi definida como 0 ou 1, 0 se o rótulo que o classificador deu à instância foi da classe negativa, e 1 caso o rótulo retornado foi da classe positiva. A diferença no uso de funções para o cálculo da probabilidade ocorreu pois as implementações das partes do experimento que envolvem esses cálculos foram feitas em momentos diferentes e não se atentou para esse fato.

Claramente em todas as classes, tabelas 3 a 7, o melhor tipo de rede em ambas as métricas, acurácia e AUC, é a *eps*. Nas classes FS, NS e NFS, o parâmetro de conversão *epsilon* teve o mesmo valor, 1,7; e o mesmo classificador, LDA, as únicas diferenças nessas classes foram o número de atributos utilizados e o tamanho da janela que resultou no melhor classificador. Importante o fato do tamanho da janela em geral ser 512 e 910 pontos, já que uma janela menor implica num menor custo computacional. Lembrando que 910 está no intervalo de tamanho de janela sugerido pela análise de estacionariedade, e que o parâmetro *epsilon* da análise de estacionariedade, que entra no cálculo da correlação integral de Grassberger-Procaccia (Eq. 3.7), foi sugerido ficar entre 1 e 2. Com exceção da classe FS, que só utilizou 4 atributos, todas as outras utilizaram 18 atributos. O fato do classificador em geral ser simples, LDA é linear, implica que o resultado tem boa generalização, e, provavelmente, mesmo utilizando menos do que os 18 atributos, o resultado ainda seria bom. Esse ótimo desempenho sinaliza uma robustez desse tipo de rede para discriminar os tipos de sinais aqui trabalhados.

Com exceção da rede *trans*, as demais redes tiveram um desempenho razoável, caso o custo computacional da rede *eps* seja demasiado para uma dada aplicação, pode-se optar pela rede *cycle* que tem apenas um parâmetro e obteve bons resultados, ou mesmo a rede *vis* que não possui parâmetros e em geral utilizou janelas menores que a *cycle*.

As redes *aknn* e *knn*, que têm um viés parecido com a rede *eps*, também tiveram bom resultado geral, embora menor que a *eps* mas em geral utilizaram menos atributos que a rede *eps*.

Conforme citado na seção 4.2.1, a análise da distribuição do número de vértices e arestas nas redes pode ajudar a entender o comportamento das medidas extraídas.

Tabela 3 – Melhores classificadores para classe FS

Função objetivo	Rede	Classificador	Atributos	Janela	Amostras	ACC	AUC teste	AUC <i>bootstrap</i>	SEN	SPC	F1
roc_auc	vis	DT - nodes: 555	17	303	507	79.88	0.19	0.81	91.03	70.33	79.01
roc_auc	aknn E0: 14	LDA -	17	1024	160	86.25	0.03	0.88	78.26	97.06	85.71
roc_auc	knn k: 9	LDA -	10	2730	40	90	0.08	0.9	86.36	94.44	89.47
roc_auc	eps epsilon_perc: 1.70	LDA -	4	2048	80	100	0	1	100	100	100
roc_auc	cycle cycle_ths: 0.20	LDA -	16	2730	40	95	0.01	0.95	94.12	95.65	95.65
roc_auc	trans bins: 100 prob_ths: 0.01	LDA -	21	910	160	55	0.5	0.5	0	100	70.97
accuracy	vis	LDA -	15	910	156	83.97	0.09	0.84	85.23	82.35	81.75
accuracy	aknn E0: 7	LDA -	9	1024	160	95.62	0.02	0.96	92.39	100	95.1
accuracy	knn k: 9	DT - nodes: 21	11	2730	40	97.5	0.02	0.98	95.65	100	97.14
accuracy	eps epsilon_perc: 1.70	LDA -	18	512	320	100	0	1	100	100	100
accuracy	cycle cycle_ths: 0.01	LDA -	16	2730	60	95	0.03	0.9	81.82	97.96	96.97
accuracy	trans bins: 100 prob_ths: 0.01	KNN - N: 3 - weights: uniform	21	2048	80	53.75	0.46	0.54	100	7.5	13.95

Tabela 4 – Melhores classificadores para classe NS

Função objetivo	Rede	Classificador	Atributos	Janela	Amostras	ACC	AUC teste	AUC <i>bootstrap</i>	SEN	SPC	F1
roc_auc	vis	DT - nodes: 107	16	1024	156	86.54	0.13	0.87	88.89	84.52	87.12
roc_auc	aknn E0: 17	LDA -	13	910	160	91.88	0.03	0.92	88.89	94.32	92.74
roc_auc	knn k: 9	LDA -	6	2730	40	87.5	0.05	0.87	84.21	90.48	88.37
roc_auc	eps epsilon_perc: 1.70	LDA -	18	910	160	100	0	1	100	100	100
roc_auc	cycle cycle_ths: 0.90	LDA -	21	2730	40	95	0.01	0.95	93.75	95.83	95.83
roc_auc	trans bins: 100 prob_ths: 0.01	LDA -	21	303	520	50	0.5	0.5	0	100	66.67
accuracy	vis	DT - nodes: 61	13	2048	78	92.31	0.07	0.93	88.1	97.22	92.11
accuracy	aknn E0: 13	DT - nodes: 31	11	2730	40	92.5	0.08	0.92	95.45	88.89	91.43
accuracy	knn k: 9	LDA -	19	1024	160	93.12	0.03	0.93	93	93.33	91.06
accuracy	eps epsilon_perc: 1.70	LDA -	18	512	320	100	0	1	100	100	100
accuracy	cycle cycle_ths: 0.50	LDA -	19	2730	40	97.5	0	0.98	100	95.65	97.78
accuracy	trans -	-	0	-	0	0	0	0	0	0	0

Tabela 5 – Melhores classificadores para classe NFS

Função objetivo	Rede	Classificador	Atributos	Janela	Amostras	ACC	AUC teste	AUC <i>bootstrap</i>	SEN	SPC	F1
roc_auc	vis	LDA -	17	2730	59	89.83	0.04	0.87	78.95	95	92.68
roc_auc	aknn E0: 17	LDA -	12	2048	120	89.17	0.06	0.85	75	95.24	92.49
roc_auc	knn k: 9	LDA -	15	1024	240	92.5	0.03	0.91	85.53	95.73	94.58
roc_auc	eps epsilon_perc: 1.70	LDA -	18	910	240	100	0	1	100	100	100
roc_auc	cycle cycle_ths: 0.70	LDA -	21	2730	60	93.33	0.02	0.88	76.47	100	95.56
roc_auc	trans bins: 100 prob_ths: 0.01	DT - nodes: 3	21	303	780	25.26	0.5	0.5	100	0.34	0.68
accuracy	vis	LDA -	15	512	472	87.92	0.07	0.84	73.03	95	91.43
accuracy	aknn E0: 17	LDA -	21	1024	240	90	0.06	0.88	82.5	93.75	92.59
accuracy	knn k: 9	DT - nodes: 153	11	1024	240	85.83	0.18	0.82	69.32	95.39	89.51
accuracy	eps epsilon_perc: 1.70	LDA -	18	512	480	100	0	1	100	100	100
accuracy	cycle cycle_ths: 0.01	LDA -	16	2730	60	95	0.03	0.9	81.82	97.96	96.97
accuracy	trans bins: 100 prob_ths: 0.01	KNN - N: 2 - weights: uniform	21	2048	120	63.33	0.51	0.5	0	100	77.55

Tabela 6 – Melhores classificadores para classe ONF

Função objetivo	Rede	Classificador	Atributos	Janela	Amostras	ACC	AUC teste	AUC <i>bootstrap</i>	SEN	SPC	F1
roc_auc	vis	SVC - C: 1 - gamma: auto - kernel: rbf - support: 211	19	910	236	94.07	0.06	0.93	91.67	95.12	95.71
roc_auc	aknn E0: 17	LDA -	9	2730	60	98.33	0.02	0.99	100	97.37	98.67
roc_auc	knn k: 12	LDA -	7	2730	59	94.92	0	0.94	93.33	95.45	96.55
roc_auc	eps epsilon_perc: 1.70	KNN - N: 7 - weights: uniform	10	2048	120	100	0	1	100	100	100
roc_auc	cycle cycle_ths: 0.10	LDA -	16	2048	120	99.17	0	0.99	100	98.89	99.44
roc_auc	trans bins: 100 prob_ths: 0.01	DT - nodes: 3	21	2048	120	30	0.49	0.51	100	2.33	4.55
accuracy	vis	LDA -	9	1024	236	99.58	0	1	100	99.29	99.64
accuracy	aknn E0: 17	LDA -	9	2048	120	92.5	0.03	0.92	90.91	93.42	94.04
accuracy	knn k: 13	LDA -	20	2730	60	91.67	0.02	0.9	85.71	94.87	93.67
accuracy	eps epsilon_perc: 1.70	KNN - N: 4 - weights: distance	18	303	780	97.56	0.02	0.96	93.16	99.45	98.28
accuracy	cycle cycle_ths: 0.50	LDA -	19	2048	120	95.83	0	0.96	95.24	96.15	96.77
accuracy	trans bins: 100 prob_ths: 0.01	KNN - N: 1 - weights: uniform	21	2730	60	65	0.5	0.5	0	100	78.79

Tabela 7 – Melhores classificadores para classe OS

Função objetivo	Rede	Classificador	Atributos	Janela	Amostras	ACC	AUC teste	AUC <i>bootstrap</i>	SEN	SPC	F1
roc_auc	vis	KNN - N: 7 - weights: distance	21	2730	39	92.31	0.02	0.93	90	94.74	92.31
roc_auc	aknn E0: 17	LDA -	13	2730	39	94.87	0	0.96	91.3	100	94.12
roc_auc	knn k: 12	LDA -	16	1024	160	95.62	0.01	0.95	94.44	96.59	96.05
roc_auc	eps epsilon_perc: 1.20	LDA -	18	256	640	100	0	1	100	100	100
roc_auc	cycle cycle_ths: 0.10	LDA -	21	2730	40	92.5	0.03	0.91	85.71	96.15	94.34
roc_auc	trans bins: 100 prob_ths: 0.01	KNN - N: 6 - weights: distance	21	512	320	47.5	0.5	0.5	0	100	64.41
accuracy	vis	LDA -	17	910	156	92.95	0.02	0.92	85.29	98.86	94.05
accuracy	aknn E0: 13	DT - nodes: 13	6	2730	40	97.5	0.03	0.97	100	94.12	96.97
accuracy	knn k: 12	LDA -	17	1024	160	95	0.02	0.95	93.18	97.22	94.59
accuracy	eps epsilon_perc: 1.20	LDA -	18	512	320	100	0	1	100	100	100
accuracy	cycle cycle_ths: 0.70	KNN - N: 3 - weights: uniform	21	2730	40	92.5	0.07	0.93	86.36	100	92.31
accuracy	trans -	- -	0	-	0	0	0	0	0	0	0

Durante a análise dos *boxplots* dos vértices e arestas, notou-se que as redes de transição apresentaram apenas um vértice e nenhuma aresta, o que sugere que há algum erro em um componente da implementação relacionado à esse algoritmo de conversão. Como essa informação foi analisada apenas no após a defesa da dissertação, preferiu-se manter os resultados para a rede de transição no texto.

O apêndice A mostra o *boxplot* das redes com os parâmetros apresentados nas tabelas 3 à 7. Um *insight* interessante dessa análise, é que apenas o número de arestas da rede de *epsilon* vizinhança parece ser suficiente para discriminar bem entre os grupos de séries analisados.

4.2.6 Comparação com a literatura

Mormann, Elger e Lehnertz (2006) listaram vários trabalhos e suas abordagens para classificação de sinais de epilepsia; apenas 4 dos 42 trabalhos listados fizeram a validação estatística. Se a validação cruzada foi utilizada nesses trabalhos apenas para treinar um modelo e estimar seu desempenho, nesse caso a estimativa reportada é confiável. Caso contrário, pode acontecer como na rede *eps* para a AUC com janela de 256 pontos, na qual o desempenho na validação cruzada foi ótimo (etapa de treino), mas ruim no conjunto de teste.

Agora com os resultados do desempenho de todas as redes e em vários cenários, é possível comparar com alguns trabalhos da literatura. A tabela 8, adaptada de duas tabelas criadas por Acharya *et al.* (2013a), reúne resultados de vários trabalhos que utilizaram a mesma base de dados deste trabalho. Diferentemente da tabela compilada por Mormann, Elger e Lehnertz (2006), a tabela 8 apresenta apenas a referência do trabalho, atributos utilizados, classificador e acurácia. Ela especifica dois cenários, um com as classes O e S, e outro com as classes O, S, e classes interictais, mas não especifica quais sinais interictais são esses, se apenas um dos dois interictais ou ambos (conjuntos N e F).

Tabela 8 – Resultados de trabalhos de detecção automática de crises. Adaptado de (ACHARYA *et al.*, 2013a).

Referência	Método	Classificador	Classes	ACC (%)
Nigam e Graupe (2004)	Filtro de processamento não linear	ANN	O,S	97.20
Kannathal <i>et al.</i> (2005b)	Medidas de entropia	ANN Fuzzy	O,S	92.22
Srinivasan <i>et al.</i> (2005)	Características de tempo e frequência	ANN recorrente	O,S	99.60
Subasi (2007)	<i>Discrete Wavelet Transform (DWT)</i>	<i>Mixture expert model</i>	O,S	94.50
Polat e Güneş (2007)	Baseado na FFT	<i>Decision tree</i>	O,S	98.72
Tzallas, Tsipouras e Fotiadis (2007)	Métodos de tempo e frequência	ANN	O,S	97.72-100
Srinivasan, Eswaran e Sridharan (2007)	ApEn	ANN probabilística	O,S	100.00
Polat e Güneş (2008a)	Baseado na FFT	Artificial immune recognition system	O,S	100.00
Polat e Güneş (2008b)	Modelo AR	C4.5 (árvore de decisão)	O,S	99.32
Ocak (2009)	DWT-ApEn	<i>Thresholding</i>	O,S	96.65
Guo <i>et al.</i> (2009)	<i>Wavelet</i>	ANN	O,S	95.20
Guo, Rivero e Pazos (2010)	ApEn e <i>Wavelet</i>	ANN	O,S	99.85
Guo <i>et al.</i> (2010)	<i>Line length</i> e <i>Wavelet</i>	ANN	O,S	99.60
Subasi e Gursay (2010)	DWT-Principal Component Analysis (PCA)	SVM	O,S	98.75
Subasi e Gursay (2010)	DWT-ICA	SVM	O,S	99.50
Subasi e Gursay (2010)	DWT-LDA	SVM	O,S	100.00
Übeyli (2010)	Modelo AR	SVM	O,S	99.56
Lima, Coelho e Eisencraft (2010)	<i>Wavelet</i>	SVM	O,S	100.00
Guo <i>et al.</i> (2011)	Algoritmos genéticos	KNN	O,S	99.00
Wang, Miao e Xie (2011)	<i>Wavelet</i> e entropia	KNN	O,S	100.00
Iscan, Dokur e Demiralp (2011)	Correlação cruzada e FFT	Vários classificadores	O,S	100.00
Orhan, Hekim e Ozer (2011)	DWT	ANN	O,S	100.00
Zhu (2012)	Distribuição do grau num grafo de visibilidade	LDA	O,S	100.00
Zhu (2012)	Distribuição do grau num grafo de visibilidade	LDA	O,F	100.00
Zhu (2012)	Distribuição do grau num grafo de visibilidade	LDA	F,S	95.00
Güler, Übeyli e Güler (2005)	Expoente de <i>Lyapunov</i>	ANN recorrente	O,interictal,S	96.79
Ghosh-dastidar e Adeli (2007)	Banda de frequência misturadas	ANN <i>spiking</i>	O,interictal,S	92.50
Ghosh-Dastidar, Adeli e Dadmehr (2007)	Banda de frequência misturadas	ANN	O,interictal,S	96.70
Ghosh-Dastidar, Adeli e Dadmehr (2008)	Banda de frequência misturadas	ANN	O,interictal,S	96.60
Ghosh-Dastidar e Adeli (2009)	Banda de frequência misturadas	ANN <i>multi-spiking</i>	O,interictal,S	90.70-94.80
Chua <i>et al.</i> (2009), Chua <i>et al.</i> (2011)	<i>Higher order spectra (HOS)</i>	GMM	O,interictal,S	93.11
Faust <i>et al.</i> (2010)	Parâmetro do domínio de frequência	SVM	O,interictal,S	93.30
Acharya <i>et al.</i> (2009)	Características não lineares	GMM	O,interictal,S	95.00
Guo <i>et al.</i> (2011)	Baseado em algoritmos genéticos	KNN	O,interictal,S	93.50
Orhan, Hekim e Ozer (2011)	DWT	ANN	O,interictal,S	96.67
Acharya, Sree e Suri (2011)	HOS dos coeficientes da <i>Wavelet Packet Decomposition (WPD)</i>	SVM	O,interictal,S	98.50
Acharya <i>et al.</i> (2011)	Parâmetros do <i>Recurrence Quantification Analysis (RQA)</i>	SVM	O,interictal,S	95.60
Acharya <i>et al.</i> (2012c)	Entropias e HOS	Fuzzy	O,interictal,S	99.70
Acharya <i>et al.</i> (2012a)	Entropias	Fuzzy	O,interictal,S	98.10
Acharya <i>et al.</i> (2012b)	PCA dos coeficientes do WPD	GMM	O,interictal,S	99.00
Acharya (2013)	DWT e ICA	SVM	O,interictal,S	96.00
Martis <i>et al.</i> (2012)	Empirical Mode	C4.5	O,interictal,S	95.33
Martis <i>et al.</i> (2012)	Empirical Mode	C4.5	O,interictal,S	95.33
Martis <i>et al.</i> (2012)	Empirical Mode	C4.5	O,interictal,S	95.33
Martis <i>et al.</i> (2012)	Empirical Mode	C4.5	O,interictal,S	95.33

A maioria dos classificadores listados na tabela 8 tem potencial para ter uma complexidade maior, e consequentemente sofrer de sobre-treino, que a maioria dos classificadores reportados nas tabelas 3 a 7. Mesmo analisando as séries temporais em janelas, que tem o efeito de criar mais de uma amostra a partir de uma série temporal, há apenas 100 séries temporais em cada um dos conjuntos (Z,O,N,F,S), e sem uma análise da complexidade dos classificadores reportados, é difícil afirmar que o resultado obtido é uma estimativa robusta do desempenho.

Feitas essas considerações, e assumindo que as estimativas de acurácia são confiáveis, é possível notar o alto desempenho das abordagens listadas na literatura. Com exceção da rede *trans*, todas as redes da classe OS têm acurácia maior que 90%, tabela 7, e há várias opções com acurácia maior que 90% nas classes NFS e ONF, vide redes cíclicas, vizinhos mais próximos, vizinhos mais próximos adaptativos e grafo de visibilidade nas tabelas 5 e 6, que são as mais próximas da parte de baixo da tabela 8. E se considerarmos a rede de *epsilon* vizinhança, ela acerta tudo exceto para a classe ONF, mas mesmo neste caso a acurácia é de 97,56% com uma janela de 303 pontos, o que é muito bom já que em geral as outras abordagens utilizam a série inteira, e um classificador de vizinhos mais próximos com k igual a 4.

A tabela 9 é relativa aos trabalhos mais recentes encontrados. Em geral, a tabela 9 apresenta resultados melhores, para o grafo de visibilidade, que os resultados apresentados nas tabelas 3, 4 e 7, as tabelas 5 e 6 contemplam cenários não utilizados nos trabalhos relacionados na tabela 9. Mas vários classificadores da tabela 9 são mais complexos que os do grafo de visibilidade nas tabelas 3, 4 e 7, e utilizam menos atributos. Sem uma análise da complexidade ou da utilização de um conjunto de teste, a confiança desses resultados é reduzida, e infelizmente esses trabalhos não mencionam se ou como esses cuidados foram tomados.

Se o foco for apenas nos classificadores LDA, os seis primeiros da segunda parte da tabela 9, fica difícil dizer que abordagem teve o melhor resultado, parecendo que há um empate entre o reportado na tabela 9 e as abordagens utilizadas nesta dissertação relativas aos grafos de visibilidade. Por outro lado, se compararmos os resultados das redes de *epsilon* vizinhança, nos cenários apresentados na tabela 9 (classes FS, NS e OS), os resultados das redes de *epsilon* vizinhança acertaram tudo utilizando um classificador linear, LDA, e em geral com tamanho de janela menor.

Tabela 9 – Compilação dos resultados do gráfico de visibilidade para os trabalhos de [Supriya et al. \(2016\)](#), [Wang et al. \(2016\)](#) e [Liu et al. \(2016\)](#).

Referência	Método	Classificador	Janela	Classes	ACC (%)	SPC (%)	SEN (%)
Supriya et al. (2016)	Grafo vis. ponderado e modularidade	SVM linear	1.024	O,S	99.5	99	100
Supriya et al. (2016)	Grafo vis. ponderado e modularidade	SVM linear	1.024	N,S	98.25	98.49	98
Supriya et al. (2016)	Grafo vis. ponderado e modularidade	SVM linear	1.024	F,S	92.5	91.66	93.33
Supriya et al. (2016)	Grafo vis. ponderado e grau	SVM linear	1.024	O,S	100	100	100
Supriya et al. (2016)	Grafo vis. ponderado e grau	SVM linear	1.024	N,S	88	100	80.64
Supriya et al. (2016)	Grafo vis. ponderado e grau	SVM linear	1.024	F,S	91.5	100	85.47
Supriya et al. (2016)	Grafo vis. ponderado grau e modularidade	SVM linear	1.024	O,S	100	100	100
Supriya et al. (2016)	Grafo vis. ponderado grau e modularidade	SVM linear	1.024	N,S	98.5	98.5	98.5
Supriya et al. (2016)	Grafo vis. ponderado grau e modularidade	SVM linear	1.024	F,S	91.25	90.24	92.3
Wang et al. (2016)	Grafo vis. ponderado	LDA	1.024	O,S	86.88	73.75	100
Wang et al. (2016)	Grafo vis. ponderado	LDA	4.097	O,S	88	76	100
Wang et al. (2016)	Grafo vis. ponderado	LDA	1.024	N,S	85.13	72	98
Wang et al. (2016)	Grafo vis. ponderado	LDA	4.097	N,S	86.5	75	98
Wang et al. (2016)	Grafo vis. ponderado	LDA	1.024	F,S	86.62	73.5	100
Wang et al. (2016)	Grafo vis. ponderado	LDA	4.097	F,S	88	76	100
Wang et al. (2016)	Grafo vis. ponderado	Q-LDA	1.024	O,S	100	100	100
Wang et al. (2016)	Grafo vis. ponderado	Q-LDA	4.097	O,S	100	100	100
Wang et al. (2016)	Grafo vis. ponderado	Q-LDA	1.024	N,S	89.62	83	96
Wang et al. (2016)	Grafo vis. ponderado	Q-LDA	4.097	N,S	89	82	96
Wang et al. (2016)	Grafo vis. ponderado	Q-LDA	1.024	F,S	98.25	98.5	98
Wang et al. (2016)	Grafo vis. ponderado	Q-LDA	4.097	F,S	98.5	99	98
Wang et al. (2016)	Grafo vis. ponderado	SVM linear	1.024	O,S	99.25	98.5	100
Wang et al. (2016)	Grafo vis. ponderado	SVM linear	4.097	O,S	95.5	91	100
Wang et al. (2016)	Grafo vis. ponderado	SVM linear	1.024	N,S	92.63	90.5	94.75
Wang et al. (2016)	Grafo vis. ponderado	SVM linear	4.097	N,S	89	82	96
Wang et al. (2016)	Grafo vis. ponderado	SVM linear	1.024	F,S	96.38	93.75	99
Wang et al. (2016)	Grafo vis. ponderado	SVM linear	4.097	F,S	92.5	85	100
Wang et al. (2016)	Grafo vis. ponderado	SVM RBF	1.024	O,S	99.12	98.25	100
Wang et al. (2016)	Grafo vis. ponderado	SVM RBF	4.097	O,S	96.5	91	100
Wang et al. (2016)	Grafo vis. ponderado	SVM RBF	1.024	N,S	93.25	91.5	95
Wang et al. (2016)	Grafo vis. ponderado	SVM RBF	4.097	N,S	91	86	96
Wang et al. (2016)	Grafo vis. ponderado	SVM RBF	1.024	F,S	97.25	95.5	99
Wang et al. (2016)	Grafo vis. ponderado	SVM RBF	4.097	F,S	95	91	99
Wang et al. (2016)	Grafo vis. ponderado	SVM Polinomial	1.024	O,S	100	100	100
Wang et al. (2016)	Grafo vis. ponderado	SVM Polinomial	4.097	O,S	100	100	100
Wang et al. (2016)	Grafo vis. ponderado	SVM Polinomial	1.024	N,S	94.5	95.25	93.75
Wang et al. (2016)	Grafo vis. ponderado	SVM Polinomial	4.097	N,S	94.5	95	94
Wang et al. (2016)	Grafo vis. ponderado	SVM Polinomial	1.024	F,S	98.38	99	97.75
Wang et al. (2016)	Grafo vis. ponderado	SVM Polinomial	4.097	F,S	98.5	98	99
Liu et al. (2016)	Grafo vis. horizontal com ApEn	SVM	2.048	F,S	89	90	88
Liu et al. (2016)	Grafo vis. horizontal com Sample Entropy	SVM	2.048	F,S	91	92	90
Liu et al. (2016)	Grafo vis. horizontal com centralidade do grau	SVM	2.048	F,S	92.98	95.98	89.98
Liu et al. (2016)	Grafo vis. horizontal com centralidade do grau	SVM	1.024	F,S	93.92	99.92	87.92

CONCLUSÃO

A hipótese de trabalho levantada na introdução foi que o uso de redes complexas permite o desenvolvimento de um algoritmo prospectivo de predição de crises epiléticas com ótimo desempenho. Para verificar essa hipótese fez-se um estudo extensivo de várias abordagens de conversão de séries temporais em redes complexas e suas métricas.

Durante o estudo sobre os métodos de conversão surgiu a questão da estacionariedade como um pré-requisito de alguns métodos, e mesmo sobre a importância dessa propriedade nos dados.

Foram utilizadas 6 abordagens de conversão, e em cada abordagem, várias configurações de parâmetros foram experimentadas resultando em 75 (8 da rede *cycle*, 1 da rede *vis*, 24 da rede *trans*, 10 da rede *eps*, 16 da rede *aknn* + 16 da rede *knn*) abordagens diferentes. Cada uma dessas abordagens foi testada em 5 subconjuntos do conjunto de dados da Universidade de Bonn, totalizando 375 cenários de conversão de série temporal num grafo. E em cada um desses cenários foram utilizados 7 tamanhos de janelas diferentes totalizando 2.625 configurações diferentes.

Com as tarefas acima descritas, o objetivo de fazer um estudo extensivo dos algoritmos de conversão é cumprido, mas devido a grande quantidade de informação produzida, os resultados individuais de cada algoritmo de conversão não foram analisados.

Cada grafo teve 21 métricas definidas na análise de redes complexas extraídas, transformando o conjunto de séries temporais num novo conjunto de dados no formato atributo-valor, em que cada instância desse novo conjunto é referente a uma janela que foi convertida da série temporal e teve suas métricas extraídas.

Nesse novo conjunto de dados foi realizada uma etapa de seleção de atributos, mas como a configuração que obteve o melhor resultado em geral utilizou quase todos os atributos isso é um indicativo que essa etapa pode ser descartada.

A fim de cumprir o segundo objetivo, comparar o desempenho dos classificadores com um preditor aleatório, foi realizada uma validação cruzada, numa parte do conjunto de dados separada para o treino, de um conjunto de 37 classificadores. Desse conjunto de classificadores, o melhor foi escolhido e seu desempenho foi avaliado num conjunto de teste. Esse desempenho foi utilizado para um teste de hipótese com um preditor aleatório, pois essa é uma crítica a várias abordagens da literatura. Os classificadores treinados foram escolhidos baseados em duas métricas, a acurácia e a AUC. A acurácia é a medida de desempenho mais reportada, e a AUC captura um balanço entre duas métricas importantes no contexto do trabalho, a sensibilidade e a especificidade. Com essas duas maneiras de selecionar o classificador, significa que foram realizados 5.250 experimentos.

Novamente, a grande quantidade de informação gerada impossibilitou uma análise individual, e por isso os resultados foram agregados em subconjuntos, as classes FS, NS, NFS, ONF e OS. Cada classe pode ser pensada como um eletrodo posicionado numa região diferente. Elas foram pensadas para avaliar o desempenho em várias situações, já que vários trabalhos que utilizam esse conjunto de dados utilizam apenas os conjuntos O (sinal de paciente saudável com olhos abertos) e S (sinal durante uma crise). Então, se o classificador tiver bom desempenho nas classes FS, por exemplo, significa que o classificador foi capaz de distinguir entre um sinal de um eletrodo capturando uma crise e outro localizado na região focal mas que ainda não está em crise.

Dentro de cada classe, os resultados foram ainda agrupados por tamanho de janela, pela métrica utilizada, acurácia ou AUC, e pelo algoritmo de conversão utilizado.

A fim de comparar o desempenho final dos classificadores, foram selecionados dentre os que se mostraram estatisticamente melhores que o preditor aleatório no conjunto de teste, um classificador por classe, por métrica, e por tipo de rede, resultando nas últimas tabelas apresentadas no final da seção de resultados 4.2.5. Da análise dessas tabelas, verificou-se que vários tipos de redes apresentaram desempenho comparável ao estado da arte, mas com a utilização de classificadores mais simples, o LDA foi o classificador mais frequente, sendo que a rede *eps* teve acurácia, sensibilidade, e especificidade de 100% no conjunto de teste em todas as classes, exceto na classe ONF cuja acurácia, sensibilidade, e especificidade foram de 97,56%, 93,16%, e 99,45%, respectivamente, e com janela de tamanho razoavelmente pequeno.

De maneira geral, classificadores simples tiveram desempenho alto o que corrobora a intuição na área de aprendizado de máquina de que os atributos são mais importante do que o classificador. No contexto do trabalho, os atributos estão fortemente relacionados a maneira que a série temporal foi convertida em grafo. Os resultados do trabalho suportam essa hipótese já que as medidas de desempenho de todas as redes foram razoavelmente consistentes; não houve casos em que um tipo de rede teve um ótimo desempenho numa configuração e em outro cenário foi péssimo.

Pode-se considerar que a abordagem utilizada no trabalho foi prospectiva pois apesar das

séries do conjunto de dados utilizadas terem sido selecionadas especificamente para representar os comportamentos de cada conjunto, obteve-se um desempenho ótimo utilizando janelas bem menores que o sinal inteiro, e sem qualquer tipo de distinção ou uso de informação *a posteriori*, e a etapa de validação corroborou as várias abordagens.

Em relação à estacionariedade, mostrou-se, baseado na literatura, a importância desta propriedade nos sinais de EEG, especialmente se os métodos de análise requerem-na. O estudo dessa propriedade deveu-se ao fato de não haver na literatura uma resposta definitiva quanto a melhor maneira de determiná-la.

Pôde-se ter uma percepção do método proposto sob duas perspectivas diferentes e complementares, e testado com dados sintéticos e reais, além da ideia dos desafios desse tipo de análise.

Dependendo dos parâmetros sendo usados, os dados podem ser vistos como estacionários como um todo, mas para uma janela diferente o sinal torna-se estacionário apenas em segmentos. Os parâmetros também são dependentes do contexto do sinal. Logo, a abordagem aqui utilizada, pode servir para segmentar o sinal facilitando a análise, e também como uma maneira de melhor entender o sinal restringindo os valores dos parâmetros na análise.

Embora esta seja uma questão antiga, ainda é importante estudá-la. O método utilizado pode ainda ser refinado já que há mais informações resultantes dele que não foram analisadas e podem servir para sua melhor avaliação.

Até o momento foi possível reduzir o intervalo dos parâmetros, o que é especialmente vantajoso no caso determinístico que é o caso mais caro computacionalmente. Percebeu-se uma sensibilidade em relação ao tamanho da janela em sinais completamente determinísticos, e a dimensão de reconstrução deve ser mantida abaixo de 5.

No caso estocástico, um nível de significância muito rígido aumenta a medida de desempenho. Mas isto não significa uma melhor discriminação da estacionariedade, significa apenas que não se pôde rejeitar a hipótese nula. A análise estocástica também apresentou sensibilidade em relação ao tamanho de janela nos dados de EEG e *Lorenz*, diferentemente dos dados Gaussianos. Isto pode ser visto como a capacidade do tamanho da janela capturar algum tipo de estrutura determinística subjacente presente nos dados.

Apesar, da abordagem de análise de séries temporais utilizada neste trabalho ter sido aplicada no contexto de epilepsia, ela pode perfeitamente ser utilizada em outros cenários como séries meteorológicas ou índices financeiros. A particularidade de cada domínio está na escolha ou ajuste dos parâmetros de conversão da série em grafo.

5.1 Limitações

Apesar de não ter sido utilizada nenhuma informação *a posteriori*, a maneira correta de aferir a prospectividade da abordagem é utilizando conjuntos de dados com duração maior. O autor tem confiança de que pelos parâmetros da rede de melhor desempenho, rede *eps*, terem sido quase que constantes nos vários cenários, porcentagem do *epsilon*, tipo de classificador, e tamanho de janela, essa configuração no conjunto de teste se mostrou prospectiva.

Outra limitação foi a não utilização de séries temporais multivariadas, que em geral é o formato que essas séries temporais são colhidas. Embora a aplicação das abordagens aqui descritas possam ser utilizadas quase que sem modificações para este tipo de dado, sua utilização não foi realizada, constituindo uma limitação do trabalho.

A falta de uma análise da complexidade algorítmica das abordagens integra outra limitação. Apesar do objetivo do trabalho ser mais uma análise extensiva quanto a adequação dos algoritmos de conversão de série temporal em grafo, para efeitos práticos, a complexidade algorítmica é uma informação importante.

5.2 Trabalhos futuros

A grande quantidade de informação gerada no trabalho abre várias frentes para trabalhos futuros.

Na questão da reconstrução do espaço de fase, utilizado tanto na análise de estacionariedade quanto nas redes *aknn*, *knn*, e *eps*, cabe a análise de valores de parâmetros consistentes em todas as janelas para a reconstrução do espaço de fase, e não como feito aqui, em que cada reconstrução foi independente em cada janela.

Como trabalho futuro, um estudo da influência dos parâmetros de conversão nos grafos em cada tipo de rede, e em como eles influenciam a distribuição das métricas extraídas também é interessante.

A análise de outras características de redes complexas como o tipo de distribuição das métricas extraídas, o ajuste a determinados tipos de redes, como WS, BA, entre outras, distribuição dos *motifs* e análise de correspondência entre o *motif* e o trecho da série temporal que deu origem, se certos *motifs* têm algum significado biológico reconhecível por especialistas. Todas são questões interessantes que podem ser expandidas desta dissertação como trabalho futuro.

Foi iniciada a implementação de uma análise temporal utilizando um horizonte de previsão, mas devido à maior complexidade e tempo demandados, focou-se na análise sem um horizonte de previsão. Como já mencionado nas limitações, o uso de conjuntos de dados de longa duração e multi canal é um trabalho futuro importante. Na análise de outras bases, é interessante

analisar se os tamanhos de janela encontrados são bons pela quantidade de informação em termos de amostras que contêm, ou se pela informação temporal contida, ou seja, se o fator crítico no desempenho não foi o tamanho da janela em si, mas o seu intervalo de tempo. Isto pode ser melhor estabelecido pegando sinais de um mesmo contexto com amostragem bem diferentes.

5.3 Contribuições

Abaixo são descritas as contribuições deste trabalho.

Uma publicação com dados resultantes deste trabalho foi apresentada na seção oral da conferência *The International Joint Conference on Neural Networks 2017* (IJCNN):

CESTARI, D. M.; ROSA, J. L. G. Stochastic and deterministic stationarity analysis of eeg data. **Proceedings of IJCNN 2017 - International Joint Conference on Neural Networks**, INNS, Anchorage, AK, USA, 2017.

Esta publicação foca apenas na análise de estacionariedade desenvolvida.

Serão preparados mais dois artigos voltados para periódicos, o primeiro expandindo os resultados de estacionariedade, e o segundo referente à classificação de sinais de epilepsia utilizando as abordagens aqui propostas.

Além das publicações como contribuição deste trabalho de mestrado, foi realizada uma parceria com o Professor Dr. Norberto Garcia Cairasco da Faculdade de Medicina de Ribeirão Preto, Departamento de Fisiologia. Dessa parceria, ainda em andamento, foi feita a reprodução de um trabalho publicado no periódico *Scientific Reports* da *Nature* sobre classificação de sinais epileptiformes. Também houve uma colaboração com o trabalho de mestrado e doutorado da Me. Polianna Delfino Pereira sobre o efeito do anticonvulsivante TMT num modelo experimental de epilepsia do lobo temporal. A convite do Professor Garcia-Cairasco o autor participou na atividade de extensão para difusão científica do evento Semana do Cérebro em duas edições, 2015 e 2016.

REFERÊNCIAS

ACHARYA, U. R. Automated diagnosis of epileptic electroencephalogram using independent component analysis and discrete wavelet transform for different electroencephalogram durations. **Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine**, SAGE Publications Sage UK: London, England, v. 227, n. 3, p. 234–244, 2013. Citado na página 126.

ACHARYA, U. R.; CHUA, C. K.; LIM, T.-C.; DORITHY; SURI, J. S. Automatic identification of epileptic EEG signals using nonlinear parameters. **Journal of Mechanics in Medicine and Biology**, World Scientific, v. 9, n. 04, p. 539–553, 2009. Citado na página 126.

ACHARYA, U. R.; MOLINARI, F.; SREE, S. V.; CHATTOPADHYAY, S.; NG, K.-H.; SURI, J. S. Automated diagnosis of epileptic EEG using entropies. **Biomedical Signal Processing and Control**, Elsevier Ltd, v. 7, n. 4, p. 401–408, jul. 2012. ISSN 17468094. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1746809411000838>>. Citado nas páginas 22, 23, 33, 34 e 126.

ACHARYA, U. R.; SREE, S. V.; ALVIN, A. P. C.; SURI, J. S. Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework. **Expert Systems with Applications**, Elsevier, v. 39, n. 10, p. 9072–9078, 2012. Citado na página 126.

ACHARYA, U. R.; SREE, S. V.; ANG, P. C. A.; YANTI, R.; SURI, J. S. Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals. **International journal of neural systems**, World Scientific, v. 22, n. 02, p. 1250002, 2012. Citado na página 126.

ACHARYA, U. R.; SREE, S. V.; CHATTOPADHYAY, S.; YU, W.; ANG, P. C. A. Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. **International journal of neural systems**, World Scientific, v. 21, n. 03, p. 199–211, 2011. Citado na página 126.

ACHARYA, U. R.; SREE, S. V.; SURI, J. S. Automatic detection of epileptic EEG signals using higher order cumulant features. **International journal of neural systems**, World Scientific, v. 21, n. 05, p. 403–414, 2011. Citado na página 126.

ACHARYA, U. R.; SREE, S. V.; SWAPNA, G.; MARTIS, R. J.; SURI, J. S. Automated EEG analysis of epilepsy: a review. **Knowledge-Based Systems**, Elsevier, v. 45, p. 147–165, 2013. Citado nas páginas 15, 125 e 126.

ACHARYA, U. R.; Vinitha Sree, S.; SWAPNA, G.; MARTIS, R. J.; SURI, J. S. Automated EEG analysis of epilepsy: A review. **Knowledge-Based Systems**, Elsevier B.V., v. 45, p. 147–165, jun. 2013. ISSN 09507051. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0950705113000798>>. Citado nas páginas 22 e 32.

ADELI, H.; GHOSH-DASTIDAR, S.; DADMEHR, N. A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. **IEEE transactions on**

bio-medical engineering, v. 54, n. 2, p. 205–11, fev. 2007. ISSN 0018-9294. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/17278577>. Citado nas páginas 33 e 34.

ADELI, H.; ZHOU, Z.; DADMEHR, N. Analysis of EEG records in an epileptic patient using wavelet transform. **Journal of neuroscience methods**, Elsevier, v. 123, n. 1, p. 69–87, 2003. Citado na página 33.

AGGARWAL, C.; SUBBIAN, K. Evolutionary Network Analysis : A Survey. **ACM Computing Surveys (CSUR)**, ACM, v. 47, n. 1, p. 1–36, 2014. Citado nas páginas 23, 24, 25 e 51.

AHADPOUR, S.; SADRA, Y.; ARASTEHFARD, Z. Markov-binary visibility graph: A new method for analyzing complex systems. **Information Sciences**, Elsevier Inc., v. 274, p. 286–302, ago. 2014. ISSN 00200255. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0020025514002679>. Citado nas páginas 24, 25 e 49.

AMINI, L.; JUTTEN, C.; ACHARD, S.; DAVID, O.; KAHANE, P.; VERCUEIL, L.; MINOTTI, L.; HOSSEIN-ZADEH, G. a.; SOLTANIAN-ZADEH, H. Comparison of five directed graph measures for identification of leading interictal epileptic regions. **Physiological measurement**, v. 31, n. 11, p. 1529–46, nov. 2010. ISSN 1361-6579. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3368828&tool=pmcentrez&rendertype=abstract>. Citado nas páginas 24 e 25.

ANDRZEJAK, R. G.; LEHNERTZ, K.; MORMANN, F.; RIEKE, C.; DAVID, P.; ELGER, C. E. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. **Physical Review E**, APS, v. 64, n. 6, p. 061907, 2001. Citado nas páginas 33, 53 e 54.

BABLOYANTZ, A.; SALAZAR, J.; NICOLIS, C. Evidence of chaotic dynamics of brain activity during the sleep cycle. **Physics Letters A**, Elsevier, v. 111, n. 3, p. 152–156, 1985. Citado na página 68.

BARRAT, A.; BARTHELEMY, M.; VESPIGNANI, A. **Dynamical processes on complex networks**. [S.l.]: Cambridge University Press Cambridge, 2008. Citado nas páginas 24, 40, 41, 42, 43 e 44.

BENDAT, J. S.; PIERSOL, A. G. **Random data: analysis and measurement procedures**. [S.l.]: John Wiley & Sons, 2011. Citado na página 38.

BERGER, H. Über das elektrenkephalogramm des menschen. **European Archives of Psychiatry and Clinical Neuroscience**, Springer, v. 87, n. 1, p. 527–570, 1929. Citado na página 29.

BERGSTROM, R. a.; CHOI, J. H.; MANDUCA, A.; SHIN, H.-S.; WORRELL, G. a.; HOWE, C. L. Automated identification of multiple seizure-related and interictal epileptiform event types in the EEG of mice. **Scientific reports**, v. 3, p. 1483, jan. 2013. ISSN 2045-2322. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3604748&tool=pmcentrez&rendertype=abstract>. Citado na página 32.

BHADURI, S.; GHOSH, D. Electroencephalographic Data Analysis With Visibility Graph Technique for Quantitative Assessment of Brain Dysfunction. **Clinical EEG and neuroscience**, p. 3–8, abr. 2014. ISSN 1550-0594. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/24781371>. Citado na página 51.

_____. Electroencephalographic data analysis with visibility graph technique for quantitative assessment of brain dysfunction. **Clinical EEG and neuroscience**, SAGE Publications Sage CA: Los Angeles, CA, v. 46, n. 3, p. 218–223, 2015. Citado na página 52.

BLANCO, S.; GARCIA, H.; QUIROGA, R. Q.; ROMANELLI, L.; ROSSO, O. Stationarity of the EEG series. **IEEE Engineering in medicine and biology Magazine**, IEEE, v. 14, n. 4, p. 395–399, 1995. Citado nas páginas 37 e 38.

BONACICH, P. Power and centrality: A family of measures. **American journal of sociology**, JSTOR, p. 1170–1182, 1987. Citado na página 44.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2013. Citado nas páginas 37, 38 e 39.

BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. **Computer networks and ISDN systems**, Elsevier, v. 30, n. 1, p. 107–117, 1998. Citado na página 44.

BULLMORE, E.; SPORNS, O. Complex brain networks: graph theoretical analysis of structural and functional systems. **Nature reviews. Neuroscience**, v. 10, n. 3, p. 186–98, mar. 2009. ISSN 1471-0048. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19190637>>. Citado nas páginas 24, 25 e 50.

CARNEY, P. R.; MYERS, S.; GEYER, J. D. Seizure prediction: methods. **Epilepsy & behavior : E&B**, Elsevier Inc., v. 22 Suppl 1, p. S94–101, dez. 2011. ISSN 1525-5069. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3233702&tool=pmcentrez&rendertype=abstract>>. Citado nas páginas 25 e 34.

CESTARI, D. M.; ROSA, J. L. G. Stochastic and deterministic stationarity analysis of eeg data. **Proceedings of IJCNN 2017 - International Joint Conference on Neural Networks**, INNS, Anchorage, AK, USA, 2017. Citado nas páginas 58 e 133.

CHANG, B. S.; LOWENSTEIN, D. H. Mechanisms of disease: Epilepsy. **New England Journal of Medicine**, v. 349, n. 13, p. 1257–1266, 2003. PMID: 14507951. Disponível em: <<http://dx.doi.org/10.1056/NEJMra022308>>. Citado nas páginas 21 e 30.

CHIU, A. W. L.; DANIEL, S.; KHOSRAVANI, H.; CARLEN, P. L.; BARDAKJIAN, B. L. Prediction of Seizure Onset in an In-Vitro Hippocampal Slice Model of Epilepsy Using Gaussian-Based and Wavelet-Based Artificial Neural Networks. **Annals of Biomedical Engineering**, v. 33, n. 6, p. 798–810, jun. 2005. ISSN 0090-6964. Disponível em: <<http://link.springer.com/10.1007/s10439-005-2346-1>>. Citado na página 34.

CHUA, K.; CHANDRAN, V.; ACHARYA, U.; LIM, C. Automatic identification of epileptic electroencephalography signals using higher-order spectra. **Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine**, SAGE Publications, v. 223, n. 4, p. 485–495, 2009. Citado na página 126.

CHUA, K. C.; CHANDRAN, V.; ACHARYA, U. R.; LIM, C. M. Application of higher order spectra to identify epileptic EEG. **Journal of Medical Systems**, Springer, v. 35, n. 6, p. 1563–1571, 2011. Citado na página 126.

CSARDI NAM LEE, T. N. G. **igraph - Network analysis software**. 2014. <<http://igraph.org/>>. Citado na página 68.

- DIKANEV, T.; SMIRNOV, D.; WENBERG, R.; VELAZQUEZ, J. P.; BEZRUCHKO, B. EEG nonstationarity during intracranially recorded seizures: statistical and dynamical analysis. **Clinical neurophysiology**, Elsevier, v. 116, n. 8, p. 1796–1807, 2005. Citado na página 58.
- DONGES, J. F.; DONNER, R. V.; KURTHS, J. Testing time series irreversibility using complex network methods. **EPL (Europhysics Letters)**, v. 102, n. 1, p. 10004, abr. 2013. ISSN 0295-5075. Disponível em: <<http://stacks.iop.org/0295-5075/102/i=1/a=10004?key=crossref.f9ec5e65bbf5ada602e2981e24ce9ebe>>. Citado nas páginas 24, 25 e 51.
- DONNER, R. V.; SMALL, M.; DONGES, J. F.; MARWAN, N.; ZOU, Y.; XIANG, R.; KURTHS, J. Recurrence-Based Time Series Analysis By Means of Complex Network Methods. **International Journal of Bifurcation and Chaos**, v. 21, n. 04, p. 1019–1046, abr. 2011. ISSN 0218-1274. Disponível em: <<http://www.worldscientific.com/doi/abs/10.1142/S0218127411029021>>. Citado nas páginas 24, 25, 46 e 47.
- ELSNER, J. B.; JAGGER, T. H.; FOGARTY, E. a. Visibility network of United States hurricanes. **Geophysical Research Letters**, v. 36, n. 16, p. L16702, ago. 2009. ISSN 0094-8276. Disponível em: <<http://doi.wiley.com/10.1029/2009GL039129>>. Citado nas páginas 24, 25 e 48.
- FAUST, O.; ACHARYA, U. R.; MIN, L. C.; SPUTH, B. H. Automatic identification of epileptic and background EEG signals using frequency domain parameters. **International journal of neural systems**, World Scientific, v. 20, n. 02, p. 159–176, 2010. Citado na página 126.
- FAWCETT, T. An introduction to roc analysis. **Pattern recognition letters**, Elsevier, v. 27, n. 8, p. 861–874, 2006. Citado na página 74.
- FERRI, R.; RUNDO, F.; BRUNI, O.; TERZANO, M. G.; STAM, C. J. The functional connectivity of different EEG bands moves towards small-world network organization during sleep. **Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology**, v. 119, n. 9, p. 2026–36, set. 2008. ISSN 1388-2457. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/18571469>>. Citado nas páginas 24 e 25.
- FRASER, A. M.; SWINNEY, H. L. Independent coordinates for strange attractors from mutual information. **Physical review A**, APS, v. 33, n. 2, p. 1134, 1986. Citado nas páginas 60 e 64.
- FREEMAN, L. C. A set of measures of centrality based on betweenness. **Sociometry**, JSTOR, p. 35–41, 1977. Citado nas páginas 42 e 43.
- FREEMAN, W. J.; BREAKSPEAR, M. Scale-free neocortical dynamics. **Scholarpedia**, v. 2, n. 2, p. 1357, 2007. Citado na página 23.
- GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. **Machine learning**, Springer, v. 63, n. 1, p. 3–42, 2006. Citado na página 70.
- GHOSH-DASTIDAR, S.; ADELI, H. Improved spiking neural networks for EEG classification and epilepsy and seizure detection. **Integrated Computer-Aided Engineering**, IOS Press, v. 14, n. 3, p. 187–212, 2007. Citado nas páginas 34 e 126.
- _____. A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection. **Neural networks**, Elsevier, v. 22, n. 10, p. 1419–1431, 2009. Citado na página 126.

GHOSH-DASTIDAR, S.; ADELI, H.; DADMEHR, N. Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection. **IEEE transactions on bio-medical engineering**, IEEE, v. 54, n. 9, p. 1545–1551, 2007. Citado na página 126.

_____. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 55, n. 2, p. 512–518, 2008. Citado na página 126.

GRASSBERGER, P.; PROCACCIA, I. Dimensions and entropies of strange attractors from a fluctuating dynamics approach. **Physica D: Nonlinear Phenomena**, Elsevier, v. 13, n. 1-2, p. 34–54, 1984. Citado na página 61.

GUERRERO-MOSQUERA, C.; TRIGUEROS, A. M.; FRANCO, J. I.; NAVIA-VÁZQUEZ, A. New feature extraction approach for epileptic EEG signal detection using time-frequency distributions. **Medical & biological engineering & computing**, v. 48, n. 4, p. 321–30, abr. 2010. ISSN 1741-0444. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20217264>>. Citado na página 32.

GÜLER, N. F.; ÜBEYLI, E. D.; GÜLER, I. Recurrent neural networks employing lyapunov exponents for EEG signals classification. **Expert systems with applications**, Elsevier, v. 29, n. 3, p. 506–514, 2005. Citado na página 126.

GUO, L.; RIVERO, D.; DORADO, J.; RABUNAL, J. R.; PAZOS, A. Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. **Journal of neuroscience methods**, v. 191, n. 1, p. 101–9, ago. 2010. ISSN 1872-678X. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20595035>>. Citado nas páginas 22, 23, 32, 33, 34 e 126.

GUO, L.; RIVERO, D.; DORADO, J.; MUNTEANU, C. R.; PAZOS, A. Automatic feature extraction using genetic programming: An application to epileptic EEG classification. **Expert Systems with Applications**, Elsevier, v. 38, n. 8, p. 10425–10436, 2011. Citado na página 126.

GUO, L.; RIVERO, D.; PAZOS, A. Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks. **Journal of neuroscience methods**, Elsevier, v. 193, n. 1, p. 156–163, 2010. Citado na página 126.

GUO, L.; RIVERO, D.; SEOANE, J. A.; PAZOS, A. Classification of EEG signals using relative wavelet energy and artificial neural networks. In: ACM. **Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation**. [S.l.], 2009. p. 177–184. Citado na página 126.

GUYON, I.; WESTON, J.; BARNHILL, S.; VAPNIK, V. Gene selection for cancer classification using support vector machines. **Machine learning**, Springer, v. 46, n. 1-3, p. 389–422, 2002. Citado na página 70.

HAO, C.; CHEN, Z.; ZHAO, Z. Analysis and prediction of epilepsy based on visibility graph. In: IEEE. **Information Science and Control Engineering (ICISCE), 2016 3rd International Conference on**. [S.l.], 2016. p. 1271–1274. Citado na página 52.

HORSTMANN, M.-T.; BIALONSKI, S.; NOENNIG, N.; MAI, H.; PRUSSEIT, J.; WELLMER, J.; HINRICHS, H.; LEHNERTZ, K. State dependent properties of epileptic brain networks: comparative graph-theoretical analyses of simultaneously recorded EEG and MEG. **Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology**, International Federation of Clinical Neurophysiology, v. 121, n. 2, p. 172–85, fev. 2010.

- ISSN 1872-8952. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20045375>>. Citado nas páginas 24, 25 e 50.
- HUBERT, L.; ARABIE, P. Comparing partitions. **Journal of classification**, Springer, v. 2, n. 1, p. 193–218, 1985. Citado na página 63.
- ISCAN, Z.; DOKUR, Z.; DEMIRALP, T. Classification of electroencephalogram signals with combined time and frequency features. **Expert Systems with Applications**, Elsevier, v. 38, n. 8, p. 10499–10505, 2011. Citado na página 126.
- KANNATHAL, N.; CHOO, M. L.; ACHARYA, U. R.; SADASIVAN, P. K. Entropies for detection of epilepsy in EEG. **Computer methods and programs in biomedicine**, v. 80, n. 3, p. 187–94, dez. 2005. ISSN 0169-2607. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/16219385>>. Citado nas páginas 32, 33 e 34.
- KANNATHAL, N.; CHOO, M. L.; ACHARYA, U. R.; SADASIVAN, P. Entropies for detection of epilepsy in EEG. **Computer methods and programs in biomedicine**, Elsevier, v. 80, n. 3, p. 187–194, 2005. Citado na página 126.
- KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: STANFORD, CA. **Ijcai**. [S.l.], 1995. v. 14, n. 2, p. 1137–1145. Citado na página 72.
- KRAMER, M. a.; CASH, S. S. Epilepsy as a disorder of cortical network organization. **The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry**, v. 18, n. 4, p. 360–72, ago. 2012. ISSN 1089-4098. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3736575&tool=pmcentrez&rendertype=abstract>>. Citado na página 31.
- LACASA, L.; LUQUE, B.; BALLESTEROS, F.; LUQUE, J.; NUÑO, J. C. From time series to complex networks : The visibility graph. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 105, n. 13, p. 4972–4975, 2008. Citado nas páginas 11, 24, 25 e 48.
- LANG, P.; LIU, D.-B.; CAI, S.-M.; HONG, L.; ZHOU, P.-L. Recurrence network analysis of the synchronous EEG time series in normal and epileptic brains. **Cell biochemistry and biophysics**, v. 66, n. 2, p. 331–6, jun. 2013. ISSN 1559-0283. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/23090787>>. Citado nas páginas 25 e 51.
- LATORA, V.; MARCHIORI, M. Efficient behavior of small-world networks. **Physical review letters**, APS, v. 87, n. 19, p. 198701, 2001. Citado na página 41.
- LEHNERTZ, K.; ANSMANN, G.; BIALONSKI, S.; DICKTEN, H.; GEIER, C.; PORZ, S. Evolving networks in the human epileptic brain. **Physica D: Nonlinear Phenomena**, Elsevier B.V., v. 267, p. 7–15, jan. 2014. ISSN 01672789. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0167278913001826>>. Citado nas páginas 23, 24 e 25.
- LEHNERTZ, K.; BIALONSKI, S.; HORSTMANN, M.-T.; KRUG, D.; ROTHKEGEL, A.; STANIEK, M.; WAGNER, T. Synchronization phenomena in human epileptic brain networks. **Journal of neuroscience methods**, v. 183, n. 1, p. 42–8, set. 2009. ISSN 1872-678X. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19481573>>. Citado na página 25.
- LEHNERTZ, K.; MORMANN, F.; OSTERHAGE, H.; MU, A.; CHERNIHOVSKIY, A.; KRUG, D.; BIALONSKI, S.; ELGER, C. E. State-of-the-Art of Seizure Prediction. v. 24, n. 2, p. 147–153, 2007. Citado nas páginas 25, 33 e 34.

LENNOX, W. G. **Science and seizures: new light on epilepsy and migraine**. [S.l.]: Harper & Brothers, 1946. Citado na página 29.

LI, P.; WANG, B.-H. Extracting hidden fluctuation patterns of Hang Seng stock index from network topologies. **Physica A: Statistical Mechanics and its Applications**, v. 378, n. 2, p. 519–526, maio 2007. ISSN 03784371. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0378437106010430>>. Citado nas páginas 24, 25 e 49.

LI, X.-G.; GAO, Z.-Y.; LI, K.-P.; ZHAO, X.-M. Relationship between microscopic dynamics in traffic flow and complexity in networks. **Physical Review E**, APS, v. 76, n. 1, p. 016110, 2007. Citado na página 49.

LIMA, C. A.; COELHO, A. L.; EISENCRAFT, M. Tackling EEG signal classification with least squares support vector machines: a sensitivity analysis study. **Computers in Biology and Medicine**, Elsevier, v. 40, n. 8, p. 705–714, 2010. Citado na página 126.

LIU, H.; MENG, Q.; ZHANG, Q.; ZHANG, Z.; WANG, D. A novel feature extraction method for epileptic seizure detection based on the degree centrality of complex network and SVM. In: SPRINGER. **International Conference on Intelligent Computing**. [S.l.], 2016. p. 170–180. Citado nas páginas 15, 52 e 128.

LORENZ, E. N. Deterministic nonperiodic flow. **Journal of the atmospheric sciences**, v. 20, n. 2, p. 130–141, 1963. Citado na página 54.

MALMIVUO, J.; PLONSEY, R. **Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields**. [S.l.]: Oxford University Press, 1995. Citado na página 29.

MANUCA, R.; SAVIT, R. Stationarity and nonstationarity in time series analysis. **Physica D: Nonlinear Phenomena**, Elsevier, v. 99, n. 2, p. 134–161, 1996. Citado nas páginas 38, 39, 59, 61, 62 e 64.

MARTIS, R. J.; ACHARYA, U. R.; TAN, J. H.; PETZNICK, A.; YANTI, R.; CHUA, C. K.; NG, E. K.; TONG, L. Application of empirical mode decomposition (EMD) for automated detection of epilepsy using EEG signals. **International journal of neural systems**, World Scientific, v. 22, n. 06, p. 1250027, 2012. Citado na página 126.

MARWAN, N.; DONGES, J. F.; ZOU, Y.; DONNER, R. V.; KURTHS, J. Complex network approach for recurrence analysis of time series. **Physics Letters A**, Elsevier B.V., v. 373, n. 46, p. 4246–4254, nov. 2009. ISSN 03759601. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0375960109011852>>. Citado nas páginas 24, 25 e 47.

MEDEIROS, D. de C.; MORAES, M. F. D. Focus on desynchronization rather than excitability: a new strategy for intraencephalic electrical stimulation. **Epilepsy & Behavior**, Elsevier, v. 38, p. 32–36, 2014. Citado na página 29.

MILO, R.; SHEN-ORR, S.; ITZKOVITZ, S.; KASHTAN, N.; CHKLOVSKII, D.; ALON, U. Network motifs: simple building blocks of complex networks. **Science**, American Association for the Advancement of Science, v. 298, n. 5594, p. 824–827, 2002. Citado na página 45.

MIROWSKI, P. W.; LECUN, Y.; MADHAVAN, D.; KUZNIECKY, R. Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. **2008 IEEE**

Workshop on Machine Learning for Signal Processing, IEEE, p. 244–249, out. 2008. Disponível em: <<http://www.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=468>>. Citado nas páginas 32 e 34.

MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. [S.l.]: Blucher, 2006. Citado nas páginas 35 e 37.

MORMANN, F. Seizure prediction. **Scholarpedia**, v. 3, n. 10, p. 5770, 2008. Citado nas páginas 11 e 65.

MORMANN, F.; ELGER, C. E.; LEHNERTZ, K. Seizure anticipation: from algorithms to clinical practice. **Current opinion in neurology**, v. 19, n. 2, p. 187–93, abr. 2006. ISSN 1350-7540. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/16538095>>. Citado nas páginas 22, 25, 32, 33, 34, 66, 74 e 125.

NEWMAN, M. E. Scientific collaboration networks. i. network construction and fundamental results. **Physical review E**, APS, v. 64, n. 1, p. 016131, 2001. Citado na página 43.

_____. Assortative mixing in networks. **Physical review letters**, APS, v. 89, n. 20, p. 208701, 2002. Citado na página 43.

_____. The structure and function of complex networks. **SIAM review**, SIAM, v. 45, n. 2, p. 167–256, 2003. Citado na página 39.

_____. **Networks: an introduction**. [S.l.]: Oxford University Press, 2010. Citado nas páginas 24, 44 e 45.

NICOLIS, G.; CANTÚ, A. G.; NICOLIS, C. Dynamical Aspects of Interaction Networks. **International Journal of Bifurcation and Chaos**, v. 15, n. 11, p. 3467–3480, nov. 2005. ISSN 0218-1274. Disponível em: <<http://www.worldscientific.com/doi/abs/10.1142/S0218127405014167>>. Citado na página 49.

NIGAM, V. P.; GRAUPE, D. A neural-network-based detection of epilepsy. **Neurological Research**, Taylor & Francis, v. 26, n. 1, p. 55–60, 2004. Citado na página 126.

NOACHTAR, S.; RÉMI, J. The role of EEG in epilepsy: a critical review. **Epilepsy & behavior : E&B**, Elsevier Inc., v. 15, n. 1, p. 22–33, maio 2009. ISSN 1525-5069. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19248841>>. Citado nas páginas 22 e 31.

NUÑEZ, A.; LACASA, L.; VALERO, E.; GÓMEZ, J. P.; LUQUE, B. Detecting Series Periodicity With Horizontal Visibility Graphs. **International Journal of Bifurcation and Chaos**, v. 22, n. 07, p. 1250160, jul. 2012. ISSN 0218-1274. Disponível em: <<http://www.worldscientific.com/doi/abs/10.1142/S021812741250160X>>. Citado nas páginas 24, 25 e 48.

NUÑEZ, A. M.; LACASA, L.; GÓMEZ, J. P.; LUQUE, B. Visibility Algorithms : A Short Review. 2012. Citado nas páginas 24, 25 e 48.

OCAK, H. Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. **Expert Systems with Applications**, Elsevier Ltd, v. 36, n. 2, p. 2027–2036, mar. 2009. ISSN 09574174. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0957417407006203>>. Citado nas páginas 33, 34 e 126.

ORHAN, U.; HEKIM, M.; OZER, M. EEG signals classification using the k-means clustering and a multilayer perceptron neural network model. **Expert Systems with Applications**, Elsevier, v. 38, n. 10, p. 13475–13481, 2011. Citado na página 126.

ORTEGA, G. J.; SOLA, R. G.; PASTOR, J. Complex network analysis of human ECoG data. **Neuroscience letters**, v. 447, n. 2-3, p. 129–33, dez. 2008. ISSN 0304-3940. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/18848970>>. Citado nas páginas 23, 24 e 25.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011. Citado nas páginas 69 e 70.

POLAT, K.; GÜNEŞ, S. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast fourier transform. **Applied Mathematics and Computation**, Elsevier, v. 187, n. 2, p. 1017–1026, 2007. Citado na página 126.

_____. Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and FFT method based new hybrid automated identification system for classification of EEG signals. **Expert Systems with Applications**, Elsevier, v. 34, n. 3, p. 2039–2048, 2008. Citado na página 126.

_____. A novel data reduction method: distance based data reduction and its application to classification of epileptiform EEG signals. **Applied Mathematics and Computation**, Elsevier, v. 200, n. 1, p. 10–27, 2008. Citado na página 126.

PONTEN, S. C.; DAFFERTSHOFER, a.; HILLEBRAND, a.; STAM, C. J. The relationship between structural and functional connectivity: graph theoretical analysis of an EEG neural mass model. **NeuroImage**, Elsevier Inc., v. 52, n. 3, p. 985–94, set. 2010. ISSN 1095-9572. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19853665>>. Citado nas páginas 24 e 25.

PRESS, W. H.; TEUKOLSKY, S. A.; VETTERLING, W. T.; FLANNERY, B. P. **Numerical recipes in C**. [S.l.]: Cambridge university press Cambridge, 1996. Citado na página 58.

REIJNEVELD, J. C.; PONTEN, S. C.; BERENDSE, H. W.; STAM, C. J. The application of graph theoretical analysis to complex networks in the brain. **Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology**, v. 118, n. 11, p. 2317–31, nov. 2007. ISSN 1388-2457. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17900977>>. Citado nas páginas 24, 25 e 49.

RUBINOV, M.; SPORNS, O. Complex network measures of brain connectivity: uses and interpretations. **NeuroImage**, Elsevier Inc., v. 52, n. 3, p. 1059–69, set. 2010. ISSN 1095-9572. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19819337>>. Citado nas páginas 23, 24 e 25.

SAERMARK, K.; LEBECH, J.; BAK, C.; SABERS, A. Magnetoencephalography and attractor dimension: normal subjects and epileptic patients. In: **Brain Dynamics**. [S.l.]: Springer, 1989. p. 149–157. Citado na página 68.

SAVIT, R.; GREEN, M. Time series and dependent variables. **Physica D: Nonlinear Phenomena**, Elsevier, v. 50, n. 1, p. 95–116, 1991. Citado na página 61.

SCHINDLER, K. a.; BIALONSKI, S.; HORSTMANN, M.-T.; ELGER, C. E.; LEHNERTZ, K. Evolving functional network properties and synchronizability during human epileptic seizures. **Chaos (Woodbury, N.Y.)**, v. 18, n. 3, p. 033119, set. 2008. ISSN 1089-7682. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19045457>>. Citado nas páginas 23, 24, 25 e 49.

SHAO, Z.-G. Network analysis of human heartbeat dynamics. **Applied Physics Letters**, v. 96, n. 7, p. 073703, 2010. ISSN 00036951. Disponível em: <<http://scitation.aip.org/content/aip/journal/apl/96/7/10.1063/1.3308505>>. Citado nas páginas 24, 25 e 48.

SHARMA, R.; PACHORI, R. B. Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. **Expert Systems with Applications**, Elsevier Ltd, v. 42, n. 3, p. 1106–1117, fev. 2015. ISSN 09574174. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0957417414005120>>. Citado nas páginas 23, 33 e 34.

SHIMADA, Y.; KIMURA, T.; IKEGUCHI, T. Analysis of Chaotic Dynamics Using Measures of. p. 61–70, 2008. Citado nas páginas 24, 25 e 47.

SHIRAZI, a. H.; Reza Jafari, G.; DAVOUDI, J.; PEINKE, J.; Reza Rahimi Tabar, M.; SAHIMI, M. Mapping stochastic processes onto complex networks. **Journal of Statistical Mechanics: Theory and Experiment**, v. 2009, n. 07, p. P07046, jul. 2009. ISSN 1742-5468. Disponível em: <<http://stacks.iop.org/1742-5468/2009/i=07/a=P07046?key=crossref.e51e19ec2790f336f82c8d2e548edeb1>>. Citado nas páginas 24, 25 e 49.

SHOEB, A. H. **Application of machine learning to epileptic seizure onset detection and treatment**. Tese (Doutorado) — Massachusetts Institute of Technology, 2009. Citado nas páginas 21, 22 e 66.

SMITH, S. J. M. EEG in the diagnosis, classification, and management of patients with epilepsy. **Journal of neurology, neurosurgery, and psychiatry**, v. 76 Suppl 2, p. ii2–7, jun. 2005. ISSN 0022-3050. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1765691&tool=pmcentrez&rendertype=abstract>>. Citado nas páginas 21 e 22.

SRINIVASAN, V.; ESWARAN, C.; SRIRAM, N. Artificial neural network based epileptic detection using time-domain and frequency-domain features. **Journal of Medical Systems**, Springer, v. 29, n. 6, p. 647–660, 2005. Citado na página 126.

SRINIVASAN, V.; ESWARAN, C.; SRIRAM, N. Approximate entropy-based epileptic EEG detection using artificial neural networks. **IEEE Transactions on information Technology in Biomedicine**, IEEE, v. 11, n. 3, p. 288–295, 2007. Citado na página 126.

STALEY, K. J.; DUDEK, F. E. Interictal spikes and epileptogenesis. **Epilepsy currents / American Epilepsy Society**, v. 6, n. 6, p. 199–202, 2006. ISSN 1535-7597. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1783497&tool=pmcentrez&rendertype=abstract>>. Citado na página 31.

STAM, C. J. Nonlinear dynamical analysis of EEG and MEG: review of an emerging field. **Clinical Neurophysiology**, Elsevier, v. 116, n. 10, p. 2266–2301, 2005. Citado na página 37.

STRAATEN, E. C. W. van; STAM, C. J. Structure out of chaos: functional brain network analysis with EEG, MEG, and functional MRI. **European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology**, Elsevier, v. 23, n. 1, p. 7–18, jan. 2013. ISSN 1873-7862. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/23158686>>. Citado nas páginas 23, 24, 25 e 51.

STROGATZ, S. H. **Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering**. [S.l.]: Westview press, 2014. Citado na página 38.

SUBASI, A. EEG signal classification using wavelet feature extraction and a mixture of expert model. **Expert Systems with Applications**, Elsevier, v. 32, n. 4, p. 1084–1093, 2007. Citado na página 126.

SUBASI, A.; GURSOY, M. I. EEG signal classification using pca, ica, lda and support vector machines. **Expert Systems with Applications**, Elsevier, v. 37, n. 12, p. 8659–8666, 2010. Citado na página 126.

SUBRAMANIAM, N. P.; HYTTINEN, J. Analysis of nonlinear dynamics of healthy and epileptic EEG signals using recurrence based complex network approach. p. 6–8, 2013. Citado nas páginas 24, 25, 50 e 51.

SUPRIYA, S.; SIULY, S.; WANG, H.; CAO, J.; ZHANG, Y. Weighted visibility graph with complex network features in the detection of epilepsy. **IEEE Access**, IEEE, v. 4, p. 6554–6566, 2016. Citado nas páginas 15, 52 e 128.

TAKENS, F. Detecting strange attractors in turbulence. In: **Dynamical systems and turbulence, Warwick 1980**. [S.l.]: Springer, 1981. p. 366–381. Citado nas páginas 38 e 60.

TEJADA, J.; COSTA, K. M.; BERTTI, P.; GARCIA-CAIRASCO, N. The epilepsies: complex challenges needing complex solutions. **Epilepsy & behavior : E&B**, Elsevier Inc., v. 26, n. 3, p. 212–28, mar. 2013. ISSN 1525-5069. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/23146364>>. Citado na página 21.

THEILER, J.; EUBANK, S. Don't bleach chaotic data. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, AIP Publishing, v. 3, n. 4, p. 771–782, 1993. Citado na página 39.

TZALLAS, A. T.; TSIPOURAS, M. G.; FOTIADIS, D. I. Automatic seizure detection based on time-frequency analysis and artificial neural networks. **Computational Intelligence and Neuroscience**, Hindawi Publishing Corporation, v. 2007, 2007. Citado na página 126.

ÜBEYLI, E. D. Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. **Expert Systems with Applications**, Elsevier, v. 37, n. 1, p. 233–239, 2010. Citado na página 126.

WANG, D.; MIAO, D.; XIE, C. Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. **Expert Systems with Applications**, Elsevier, v. 38, n. 11, p. 14314–14320, 2011. Citado na página 126.

WANG, H.; ZHUO, G.; ZHANG, Y. *et al.* Analyzing EEG signal data for detection of epileptic seizure: Introducing weight on visibility graph with complex network feature. In: **SPRINGER. Australasian Database Conference**. [S.l.], 2016. p. 56–66. Citado nas páginas 15, 52 e 128.

WHO. **Epilepsy**. 2017. [Online; accessed 13-March-2017]. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs999/en/>>. Citado na página 21.

WIKIPEDIA. **10-20 system (EEG)** — **Wikipedia, The Free Encyclopedia**. 2016. [Online; accessed 10-November-2016]. Disponível em: <[https://en.wikipedia.org/w/index.php?title=10-20_system_\(EEG\)&oldid=7488090](https://en.wikipedia.org/w/index.php?title=10-20_system_(EEG)&oldid=7488090)>. Citado nas páginas 11 e 30.

WITT, A.; KURTHS, J.; PIKOVSKY, A. Testing stationarity in time series. **physical Review E**, APS, v. 58, n. 2, p. 1800, 1998. Citado na página 58.

XU, X.; ZHANG, J.; SMALL, M. Superfamily phenomena and motifs of networks induced from time series. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 50, p. 19601–5, dez. 2008. ISSN 1091-6490. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2604928&tool=pmcentrez&rendertype=abstract>>. Citado nas páginas 24, 25 e 47.

YANG, B.; LIU, J.; LIU, D. Characterizing and extracting multiplex patterns in complex networks. **IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics**, v. 42, n. 2, p. 469–481, 2012. Citado nas páginas 24 e 25.

YANG, Y.; YANG, H. Complex network-based time series analysis. **Physica A: Statistical Mechanics and its Applications**, v. 387, n. 5-6, p. 1381–1386, fev. 2008. ISSN 03784371. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0378437107011235>>. Citado nas páginas 24, 25 e 47.

YU, X.; LIU, C.; WANG, J.; DAI, J.; LI, J.; HOU, F. Electroencephalogram signal analysis based on the improved k-nearest neighbor network. In: IEEE. **Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), International Congress on**. [S.l.], 2016. p. 1492–1497. Citado na página 52.

ZHANG, J.; SMALL, M. Complex Network from Pseudoperiodic Time Series: Topology versus Dynamics. **Physical Review Letters**, v. 96, n. 23, p. 238701, jun. 2006. ISSN 0031-9007. Disponível em: <<http://link.aps.org/doi/10.1103/PhysRevLett.96.238701>>. Citado nas páginas 24, 25 e 46.

ZHANG, J.; SUN, J.; LUO, X.; ZHANG, K.; NAKAMURA, T.; SMALL, M. Characterizing pseudoperiodic time series through the complex network approach. **Physica D: Nonlinear Phenomena**, Elsevier B.V., v. 237, n. 22, p. 2856–2865, nov. 2008. ISSN 01672789. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0167278908001930>>. Citado nas páginas 24, 25 e 47.

ZHU, G. Analysing epileptic EEGs with a visibility graph algorithm. **2012 5th International Conference on BioMedical Engineering and Informatics**, IEEE, n. Bmei, p. 432–436, out. 2012. Disponível em: <<http://www.ieeeexplore.org/lpdocs/epic03/wrapper.htm?arnumber=651>>. Citado nas páginas 24, 25, 50 e 126.

ZHU, G.; LI, Y.; WEN, P. P. Epileptic seizure detection in EEGs signals using a fast weighted horizontal visibility algorithm. **Computer methods and programs in biomedicine**, Elsevier Ireland Ltd, v. 115, n. 2, p. 64–75, jul. 2014. ISSN 1872-7565. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24768081>>. Citado nas páginas 11, 23, 24, 25, 32 e 51.

ZOLDI, S. M.; KRYSTAL, A.; GREENSIDE, H. S. Stationarity and redundancy of multichannel EEG data recorded during generalized tonic-clonic seizures. **Brain topography**, Springer, v. 12, n. 3, p. 187–200, 2000. Citado nas páginas 37 e 58.

ANÁLISE DE VÉRTICES E ARESTAS

Como mencionado na seção 4.2.1, este apêndice traz mais detalhes sobre a distribuição do número de vértices e arestas para cada tipo de algoritmo de conversão e cada parametrização utilizada. Aqui são relacionados os *boxplots* das configurações que resultaram nos melhores classificadores dentre os vários cenários analisados, vide tabelas 3 a 7. Sempre na primeira figura são apresentados os *boxplots* do número de vértices para uma dada configuração do algoritmo de conversão para todos os tamanho de janelas utilizados, e na figura seguinte é apresentado o *boxplot* do número de arestas considerando o mesmo princípio.

As figuras 50 a 57 mostram os *boxplots* da rede de vizinhos mais próximos adaptativos para o parâmetro $E0$ igual a 7, 13, 14 e 17. É possível ver que a variação do parâmetro $E0$ não afeta a distribuição do número de vértices para cada tamanho de janela utilizada, e como esperado, uma janela maior reflete num maior número de vértices. Quanto ao número de arestas, é interessante que produz muitos *outliers* e com uma distribuição maior nas janelas de tamanho até 512 pontos. Diferentemente do que ocorre com os vértices, o número de arestas parece aumentar um pouco conforme $E0$ aumenta, para um mesmo tamanho de janela.

Figura 50 – *Boxplot* do número de vértices da rede de vizinhos mais próximos adaptativos com parâmetro $E0$ igual a 7.

Figura 51 – *Boxplot* do número de arestas da rede de vizinhos mais próximos adaptativos com parâmetro $E0$ igual a 7.

Figura 52 – *Boxplot* do número de vértices da rede de vizinhos mais próximos adaptativos com parâmetro $E0$ igual a 13.

Figura 53 – *Boxplot* do número de arestas da rede de vizinhos mais próximos adaptativos com parâmetro $E0$ igual a 13.

Figura 54 – *Boxplot* do número de vértices da rede de vizinhos mais próximos adaptativos com parâmetro $E0$ igual a 14.

Figura 55 – Boxplot do número de arestas da rede de vizinhos mais próximos adaptativos com parâmetro $E0$ igual a 14.

Figura 56 – *Boxplot* do número de vértices da rede de vizinhos mais próximos adaptativos com parâmetro $E0$ igual a 17.

Figura 57 – Boxplot do número de arestas da rede de vizinhos mais próximos adaptativos com parâmetro $E0$ igual a 17.

As figuras 58 a 63 mostram os *boxplots* da rede de vizinhos mais próximos para o parâmetro k igual a 9, 12 e 13. Nestas redes, tanto o número de vértices quanto o de arestas com janelas até 512, apresentam bastante *outliers*. Para dado valor de k , o número de vértices tem um aumento com o aumento da janela, mas a distribuição entre janelas vizinhas não parece sofrer muita variação, apesar do aumento de vértices para janelas maiores. Como esperado, o número de arestas aumento com o aumento da janela utilizada, fixado o valor do parâmetro. Interessante, que o parâmetro k igual a 12 apresentou o maior número de arestas. Novamente, há grande presença de *outliers* para janelas de tamanho até 512.

Figura 58 – *Boxplot* do número de vértices da rede de vizinhos mais próximos com parâmetro k igual a 9.

Figura 59 – *Boxplot* do número de arestas da rede de vizinhos mais próximos com parâmetro k igual a 9.

Figura 60 – *Boxplot* do número de vértices da rede de vizinhos mais próximos com parâmetro k igual a 12.

Figura 61 – *Boxplot* do número de arestas da rede de vizinhos mais próximos com parâmetro k igual a 12.

Figura 62 – *Boxplot* do número de vértices da rede de vizinhos mais próximos com parâmetro k igual a 13.

Figura 63 – *Boxplot* do número de arestas da rede de vizinhos mais próximos com parâmetro k igual a 13.

As figuras 64 a 67 mostram os *boxplots* da rede de *epsilon* vizinhança para o parâmetro *epsilon* igual a 1,2 e 1,7. Diferente das redes anteriores, o número de vértices para as séries de uma crise, apresentou uma grande dispersão para os dois parâmetros utilizados em *epsilon* e para todos os tamanhos de janelas. Nos outros casos, o comportamento dos vértices foi como o das redes anteriores. Por outro lado, o comportamento das arestas foi bem diferente. As séries F e N, tiveram dispersão parecida para os dois valores de *epsilon* reportados, sendo que para *epsilon* igual a 1,2 a dispersão foi alta, e para *epsilon* de 1,7 em geral foi uma dispersão baixa. E o comportamento das séries do tipo O foi o oposto das séries do tipo F e N. As séries do tipo S, crise, tiveram uma dispersão bem mais baixa que as outras séries. Esse comportamento das arestas é a provável razão do bom desempenho deste tipo de rede. Apenas o número de arestas parece ser suficiente para discriminar bem entre os grupos de séries analisados.

Figura 64 – Boxplot do número de vértices da rede de ϵ vizinhança com parâmetro ϵ igual a 1,2.

Figura 65 – *Boxplot* do número de arestas da rede de *epsilon* vizinhança com parâmetro *epsilon* igual a 1,2.

Figura 66 – Boxplot do número de vértices da rede de ϵ vizinhança com parâmetro ϵ igual a 1,7.

Figura 67 – *Boxplot* do número de arestas da rede de ϵ vizinhança com parâmetro ϵ igual a 1,7.

As figuras 68 a 79 mostram os *boxplots* da rede cíclica para o parâmetro *limiar de correlação* igual a 0,01; 0,1; 0,2; 0,5; 0,7 e 0,9. Tanto a distribuição dos vértices quanto a de arestas deste tipo de rede parece ser invariante ao valor do parâmetro utilizado, há apenas um aumento no número de vértices e arestas conforme aumenta-se o tamanho da janela. Mas o posicionamento das distribuições para os tipos de séries se mantém quase constantes.

Figura 68 – *Boxplot* do número de vértices da rede cíclica com parâmetro *limiar de correlação* igual a 0,01.

Figura 69 – *Boxplot* do número de arestas da rede cíclica com parâmetro *limiar de correlação* igual a 0,01.

Figura 70 – Boxplot do número de vértices da rede cíclica com parâmetro *limiar de correlação* igual a 0,1.

Figura 71 – *Boxplot* do número de arestas da rede cíclica com parâmetro *limiar de correlação* igual a 0,1.

Figura 72 – Boxplot do número de vértices da rede cíclica com parâmetro *limiar de correlação* igual a 0,2.

Figura 73 – *Boxplot* do número de arestas da rede cíclica com parâmetro *limiar de correlação* igual a 0,2.

Figura 74 – Boxplot do número de vértices da rede cíclica com parâmetro *limiar de correlação* igual a 0,5.

Figura 75 – *Boxplot* do número de arestas da rede cíclica com parâmetro *limiar de correlação* igual a 0,5.

Figura 76 – Boxplot do número de vértices da rede cíclica com parâmetro *limiar de correlação* igual a 0,7.

Figura 77 – *Boxplot* do número de arestas da rede cíclica com parâmetro *limiar de correlação* igual a 0,7.

Figura 78 – Boxplot do número de vértices da rede cíclica com parâmetro *limiar de correlação* igual a 0,9.

Figura 79 – *Boxplot* do número de arestas da rede cíclica com parâmetro *limiar de correlação* igual a 0,9.

As figuras 80 a 81 mostram os *boxplots* do grafo de vizinhança, esse algoritmo não possui parâmetro de conversão. Como nesse tipo de grafo cada amostra da série é transformada num vértice a distribuição dos vértices acompanha o tamanho da janela. O número de arestas por outro lado tem um aumento expressivo conforme aumenta-se o tamanho da janela, mas a distribuição relativa entre os tipos de séries mantém-se quase constante.

Figura 80 – *Boxplot* do número de vértices do grafo de vizinhança.

Os *boxplots* relativos às redes de transição não foram reportados pois todos apresentaram valor zero, o que indica algum erro no processo de conversão utilizando este algoritmo.

Figura 81 – *Boxplot* do número de arestas do grafo de vizinhança.

EXTRAÇÃO DAS MÉTRICAS

No *boxplot* de cada medida, o eixo x representa o tamanho de janela utilizado na conversão. Os valores de 2 a 8 representam as janelas de 256, 303, 512, 910, 1.024, 2.048, e 2.730 pontos, e o valor 1 refere-se ao *boxplot* com todas as janelas juntas. Nenhum registro foi removido para a geração desses gráficos, isso explica o porque para algumas medidas, não há valores em algumas janelas. Provavelmente, para esses valores de janelas, há registros com valores inválidos, fazendo com que o *boxplot* não seja gerado.

Com esses *boxplots*, é possível ter uma ideia geral de como os algoritmos, e também o tamanho da janela, influenciam as medidas extraídas. Um dos aspectos mais visíveis é a mudança na escala de uma mesma medida para os vários algoritmos e suas janelas. A complexidade (*compl*), por exemplo, tem uma variação entre as janelas e entre os algoritmos de conversão; apenas a rede *aknn* manteve os valores razoavelmente constantes para as várias janelas. Segundo os *boxplots*, é possível ver que alguns algoritmos tem propensão de produzir *outliers* na maioria, ou em todas as métricas extraídas da rede; *vis* e *cycle* são exemplos deste comportamento. Outro fato que merece atenção é a razoável constância de algumas métricas em alguns algoritmos, na rede *aknn* há várias métricas com uniformidade em relação ao tamanho da janela. Essa uniformidade, provavelmente, significa que a característica medida pela métrica em questão não é afetada pela janela.

Muitas outras conclusões podem ser extraídas da análise desses *boxplots* considerando apenas os algoritmos de conversão e os tamanhos de janelas utilizadas, mas como o foco do trabalho está na classificação, segue-se para a próxima etapa do fluxo de trabalho, o pré-processamento.

Figura 82 – Boxplot da medidas da rede de vizinhos mais próximos adaptativos.

Figura 83 – Boxplot da medidas da rede de vizinhos mais próximos.

Figura 84 – Boxplot da medidas da rede de ϵ vizinhança.

Figura 85 – Boxplot da medidas da rede cíclica.

Figura 86 – Boxplot da medidas da rede de transição.

Figura 87 – Boxplot da medidas do grafo de visibilidade.

SELEÇÃO DE ATRIBUTOS

Abaixo, são mostrados gráficos de barra, figura 88, com a contagem de cada atributo na seleção final de atributos. Essa seleção é feita para cada parâmetro de conversão utilizado, por isso os valores totais dos gráficos são diferentes. Eles dependem da quantidade de conversões realizadas em cada algoritmo de conversão. Por exemplo, a conversão de redes cíclicas tem dois parâmetros, o número de intervalos (*bins*) e o limiar de probabilidade, resultando numa combinação de 24 parâmetros diferentes (4 valores para intervalos e 6 valores de limiares), combinado aos parâmetros, 5 classes de séries temporais diferentes (NS, FS, NFS, ONF e OS), juntamente com as duas métricas utilizadas (acurácia e AUC), e os 7 tamanhos diferentes de janelas utilizados, são realizadas 1.680 conversões. Portanto, se algum atributo fosse selecionado em todas as conversões, ele teria uma contagem em seu respectivo gráfico de barras de 1.680 pontos.

As classes de séries temporais diferentes são os cenários explicados no final da seção 3.2.2.5, que visam analisar sinais de situações que misturam período de crise, período fora da crise, e sinal de paciente saudável, e a classe identifica qual é a situação. Por exemplo, a classe NS contém sinal de paciente com epilepsia de um eletrodo fora da região de foco da crise (N), e sinal durante a crise (S). Novamente, como a análise individual para cada parâmetro ou conjunto de parâmetros é impraticável devido à grande quantidade de combinações, vide tabela 10, são mostrados apenas os gráficos de cada algoritmo de conversão considerando todas as combinações realizadas.

A tabela 10 mostra um exemplo do resultado da seleção de atributos do grafo de visibilidade com séries dos conjuntos N e S. A tabela é dividida verticalmente em duas, uma para cada métrica utilizada no processo de seleção, acurácia a direita, e AUC a esquerda. Cada métrica está subdividida em vários tamanhos de janelas. Cada célula pintada de cinza indica se o atributo correspondente foi selecionado para a métrica e tamanho de janela em questão. Interessante notar no exemplo, a influência do tamanho da janela no processo de seleção de atributos. A janela

Tabela 10 – Seleção de atributos do grafo de visibilidade da classe NS

métrica	ACC							AUC						
	256	303	512	910	1.024	2.048	2.730	256	303	512	910	1.024	2.048	2.730
compl														
trans														
cpd														
pagerank_mean														
pagerank_var														
betweenness_mean														
betweenness_var														
coreness_mean														
coreness_var														
eigenvector_mean														
eigenvector_var														
knn_mean														
knn_var														
dg_entropy														
clus_coef_mean														
clus_coef_var														
closeness_mean														
closeness_var														
degree_mean														
degree_var														
assort														

de tamanho 512, em ambas as métricas, selecionou apenas 4 atributos. Outro fato curioso é a seleção recorrente de alguns atributos independente do tamanho da janela ou mesmo da métrica utilizada.

Os quatro atributos mais frequentemente selecionados na rede *aknn* são: média do *PageRank*, variância do coeficiente de aglomeração, média da centralidade por proximidade, e média do coeficiente de aglomeração. No geral, os atributos tiveram uma variação grande entre si, fato menos pronunciado do ocorrido na rede *knn*, cujo viés de transformação da série temporal em grafo é parecido com a *aknn*. Na rede *knn* a média de intermediação também foi selecionada várias vezes, mas as métricas, média do *PageRank*, média da centralidade por proximidade, e variância do coeficiente de aglomeração, foram as menos selecionadas. Os quatro atributos mais selecionados nesse algoritmo foram variância do *PageRank*, complexidade, média da centralidade por *K-core*, e dominância do ponto central. Nesse tipo de rede, a variação geral entre os atributos foi menor se comparada à da rede *aknn*.

A rede *eps* (*epsilon* vizinhança), não teve as 21 métricas grafadas; as três que estão faltando, assortatividade, média do grau, e variância do grau. A assortatividade não foi selecionada, provavelmente, porque deve ter tido boa parte de seus valores como inválidos, já o porque da média e variância do grau não terem sido selecionadas não está claro. Talvez, não o foram pelo simples fato de não serem discriminativas para a classe da instância. A rede *eps* teve uma baixa variação na contagem dos atributos selecionados, e a variância da centralidade por *K-core* foi o atributos mais selecionado, seguida da média da intermediação, variância do coeficiente de aglomeração, e entropia do grau.

As três redes descritas acima têm o mesmo viés, e calculam algum tipo de medida de

Figura 88 – Contagem de cada atributo após a seleção de atributos.

proximidade no espaço de fase da série temporal, enquanto que as próximas três têm vieses completamente diferentes. A rede *cycle* apresenta alguma variação na seleção dos atributos, mas menor que a da rede *aknn*, e os atributos mais selecionados foram a variância da intermediação, variância do *PageRank*, transitividade, e variância do grau.

A rede *trans* apresentou o comportamento mais diferente, todos os atributos parecem ter

a mesma contagem, embora esse não seja o caso, pois selecionando alguns valores de janelas ou classes, uma pequena diferença na seleção dos atributos é distinguível.

Por fim, na seleção da rede *vis*, é possível ver que ocorre o menor número de conversões, a variação entre os atributos parece com a da rede *knn*. Seus atributos mais selecionados foram a média do coeficiente de aglomeração, média da intermediação, variância do coeficiente de aglomeração, e a transitividade.