Camila Sardeto Deolindo

Identificação de assembleias neurais em trens de *spike*: métricas de similaridade e métodos de classificação

Macaíba

2015

INSTITUTO INTERNACIONAL
DE NEUROCIÊNCIAS
EDMOND E LILY SAFRA
Programa de Pós-Graduação em Neuroengenharia

Camila Sardeto Deolindo

Identificação de assembleias neurais em trens de *spike*: métricas de similaridade e métodos de classificação

Dissertação apresentada ao Programa de Pós-Graduação em Neuroengenharia do Instituto Internacional de Neurociências - Edmond e Lily Safra como parte dos requisitos exigidos para a obtenção do título de Mestre em Neuroengenharia.

Orientador: Renan Cipriano Moioli, Dr.

Co-orientador Fabrício Lima Brasil, Dr.

Macaíba

2015

Camila Sardeto Deolindo

Identificação de assembleias neurais em trens de *spike*: métricas de similaridade e métodos de classificação

Dissertação apresentada ao Programa de Pós-Graduação em Neuroengenharia do Instituto Internacional de Neurociências - Edmond e Lily Safra como parte dos requisitos exigidos para a obtenção do título de Mestre em Neuroengenharia.

Trabalho aprovado. Macaíba, 14 de Agosto de 2015:

Renan Cipriano Moioli, Dr.
Orientador

Ana Carolina Bione Kunicki, Dr.
Membro avaliador interno

Anne Magaly de Paula Canuto, Dr.
Membro avaliador externo

Macaíba
2015

Resumo

A caracterização funcional de assembleias neurais é um grande desafio da neurociência. No paradigma atual, a análise dos trens de *spike* inicia-se com a proposta de uma métrica de similaridade que quantifica a relação entre a atividade elétrica dos neurônios registrados. Algoritmos de agrupamento são aplicados a fim de identificar quais desses neurônios formam cada assembleia. Há diversas abordagens para o problema, cada uma utilizando uma combinação específica de métodos de processamento e atributos da atividade neural (taxa média de disparos, intervalo entre disparos, padrões temporais, entre outros). No entanto, não há consenso sobre qual abordagem retorna os melhores resultados. Desta forma, neste projeto executamos a análise comparativa de diversas combinações de métricas de similaridade e algoritmos de agrupamento para a caracterização de assembleias neurais. Em especial, considerando-se que a maioria dos métodos existentes pressupõe interações puramente excitatórias entre neurônios, o projeto estende as análises para cenários com inibição e com atrasos, dada a relevância dessas interações na atividade neuronal e ao fato delas serem pouco exploradas por modelos computacionais. Foram utilizados dados artificiais, criados a partir de modelos biologicamente plausíveis, mais especificamente modelos com base em distribuições de Poisson e modelos gaussianos latentes multivariados. Adicionalmente, analisamos dados obtidos a partir do registro multiunitário em ratos *Long-Evans* executando uma tarefa de discriminação tátil. Os resultados contribuem para fundamentar a escolha de métodos de processamento de sinais para identificação de assembleias neurais, bem como para delinear as limitações de cada abordagem. As descobertas são relevantes principalmente para o projeto de novas interfaces cérebro-máquina e para o estudo do código neural.

Palavras-chaves: Assembleias neurais 1; Reconhecimento de padrões 2; Simulação neural 3.

Abstract

The functional characterization of neuronal assemblies is a major challenge in neuroscience. In the actual framework, the spike train analysis starts with the quantification of neuronal interactions, through the definition of a similarity metric. A pattern detection algorithm that aims to identify the cells composing each assembly. Several approaches have been proposed, each employing a specific composition of algorithm and features of neuronal activity (mean firing rate, inter spike interval,temporal patterns, among others). However, it is not known which approach leads to best results. Thus, in this project we performed a comparison among different similarity measures and clustering algorithms specific to the problem of neuronal assemblies characterization. Considering that the majority of the existing methodologies presume solely excitatory interactions between neurons, we extended the analysis to scenarios that also take inhibitions and time-delayed interactions into account, given the relevance of these interactions in neuronal activity and the fact that they are little explored by computer models. Artificial data were simulated, from biologically plausible models, more specifically models based on Poisson distributions and latent multivariate Gaussian models. In addition, we processed data obtained from multiunit records in Long-Evans rats performing a tactile discrimination task. The results contribute to fundament the choice of signal processing methods for neuronal assemblies characterization, as well as to delineate the limitations of each approach. The findings are relevant for the design of new brain-machine interfaces and for the understanding of the neural code.

Keywords: Neuronal Assemblies 1; Pattern detection 2; Neural Simulation 3.

Sumário

1 Introdução 1

1.1 Assembleias Neurais 1

1.2 Trens de *spike*: Propriedades e terminologias 4

1.3 O código neural 6

1.4 Assembleias neurais em trens de *spike* 9

1.5 Objetivos da dissertação 10

2 Métodos 13

2.1 Geração de trens de *spike* artificiais 13

2.1.1 Seleção de características de dados e populações reais. 13

2.1.2 Geração de trens de *spike* a partir da distribuição de Poisson 15

2.1.3 Geração de trens de *spike* a partir da correlação entre neurônios . . . 15

2.2 Métricas de similaridade entre trens de *spike* 17

2.2.1 Métricas de similaridade baseadas na taxa de disparos 17

2.2.1.1 Correlação 18

2.2.1.2 Hamming 18

2.2.2 Métricas de similaridade com escala temporal variável 18

2.2.2.1 Victor e Purpura 20

2.2.2.2 Convolução com uma janela e Cosseno 21

2.2.2.3 Van Rossum 21

2.2.2.4 Billeh 22

2.2.3 Métricas de similaridade baseadas nas relações temporais entre *spikes* 23

2.2.3.1 Sincronização de evento 23

2.2.3.2 Intervalo entre *spikes* (ISI) 24

2.2.3.3 SPIKE 25

2.2.4 Adaptações 27

2.2.5 Comentário geral sobre as limitações das métricas de similaridade . . 28

2.3 Métodos de classificação 29

2.3.1 Análise de componentes principais (PCA) e Decomposição em valores singulares (SVD) 30

2.3.2 Análise de componentes independentes (ICA) 32

2.3.3 Modularidade 36

2.3.4 Análise do agrupamento hierárquico 39

2.3.4.1 Ligação Simples 40

2.3.4.2 Ligação Completa . . . 42
2.3.4.3 Ligação Média . . . 42
2.3.4.4 Índices de validação . . . 45
2.3.4.4.1 Índice de Davies-Bouldin . . . 45
2.3.4.4.2 Índice Silhouette . . . 46
2.3.5 Estabilidade de uma partição . . . 47
2.3.6 Adaptações . . . 49
2.3.7 Comentário geral sobre as limitações dos métodos de classificação . . 50
2.4 Métricas de desempenho . . . 50
2.4.1 Número de assembleias detectadas . . . 51
2.4.2 Classificação dos neurônios . . . 51
2.4.3 Exemplo . . . 52
2.5 Descrição do experimento em modelo animal . . . 53
3 Resultados . . . 57
3.1 Simulações . . . 57
3.1.1 Descrição das ANs simuladas . . . 57
3.1.2 Resultados: ICA . . . 60
3.1.2.1 ICA: Simulações de ANs independentes . . . 60
3.1.2.1.1 ICA: Simulações utilizando trens de *spike* segundo a distribuição de Poisson em ANs independentes . . . 60
3.1.2.1.2 ICA: Simulações utilizando trens de *spike* que consideram a correlação entre neurônios em ANs independentes . . . 65
3.1.2.2 ICA: Simulações de ANs com ao menos um neurônio exclusivo 68
3.1.2.2.1 ICA: Simulações utilizando trens de *spike* segundo a distribuição de Poisson em ANs com ao menos um neurônio exclusivo . . . 68
3.1.2.2.2 ICA: Simulações utilizando trens de *spike* que consideram a correlação entre neurônios em ANs com ao menos um neurônio exclusivo . . . 70
3.1.2.3 ICA: Simulações de ANs sem neurônios disjuntos . . . 72
3.1.2.3.1 ICA: Simulações utilizando trens de *spike* segundo a distribuição de Poisson em ANs sem neurônios disjuntos . . . 72
3.1.2.3.2 ICA: Simulações utilizando trens de *spike* que consideram a correlação entre neurônios em ANs sem neurônios disjuntos . . . 76

3.1.3 Resultados: Modularidade . . . 80
3.1.3.1 Modularidade: Simulações de ANs independentes . . . 80
3.1.3.1.1 Modularidade: Simulações utilizando trens de *spike* segundo a distribuição de Poisson em ANs independentes . . . 80
3.1.3.2 Modularidade: Simulações de ANs que compartilham neurônios 85
3.1.4 Resultados: Agrupamento Hierárquico . . . 86
3.1.4.0.1 Agrupamento Hierárquico: Simulações utilizando trens de *spike* segundo a distribuição de Poisson em ANs independentes . . . 86
3.1.4.0.2 Agrupamento Hierárquico: Simulações utilizando trens de *spike* que consideram a correlação entre neurônios em ANs sem neurônios disjuntos . . . 91
3.1.5 Resultados: Estabilidade de uma partição . . . 97
3.1.5.1 Estabilidade de uma partição: Simulações de ANs Independentes . . . 97
3.1.5.2 Estabilidade de uma partição: Simulações de ANs que compartilham neurônios . . . 100
3.2 Resultados da aplicação dos algoritmos a registros eletrofisiológicos extracelulares 102
4 Discussão . . . 109
4.1 Propriedades e limitações das métricas e algoritmos estudados . . . 109
4.1.1 Métricas de similaridade . . . 109
4.1.2 Métodos de classificação . . . 114
4.2 Discussão dos resultados para conjuntos de dados simulados . . . 117
4.2.1 Desempenho dos algoritmos com diferentes taxas médias de disparo . 118
4.2.1.1 Discussão: Variação do desempenho do ICA com diferentes taxas médias de disparo . . . 118
4.2.1.2 Discussão: Variação do desempenho da Modularidade com diferentes taxas médias de disparo . . . 119
4.2.1.3 Discussão: Variação do desempenho do Agrupamento Hierárquico com diferentes taxas médias de disparo . . . 120
4.2.1.4 Discussão: Variação do desempenho da Estabilidade de uma partição com diferentes taxas médias de disparo . . . 121
4.2.2 Desempenho dos algoritmos com ANs de tamanhos diferentes . . . 122
4.2.2.1 Discussão: Variação do desempenho do ICA com ANs de tamanhos diferentes . . . 122

4.2.2.2 Discussão: Variação do desempenho da Modularidade com ANs de tamanhos diferentes . . . 122
4.2.2.3 Discussão: Variação do desempenho do Agrupamento Hierárquico com ANs de tamanhos diferentes . . . 123
4.2.2.4 Discussão: Variação do desempenho da Estabilidade de uma partição com ANs de tamanhos diferentes . . . 123
4.2.3 Desempenho dos algoritmos em ANs considerando diferentes disposições dos neurônios . . . 124
4.2.4 Desempenho dos algoritmos em interações com atraso de fase . . . 125
4.2.5 Desempenho dos algoritmos segundo um modelo de disparos mais realista 126
4.2.6 Sumário dos Resultados . . . 127
4.3 Discussão dos resultados referentes à aplicação dos algoritmos a registros eletrofisiológicos extracelulares . . . 129
5 Conclusões . . . 131
5.1 Perspectivas e propostas de extensões . . . 132
5.2 Consideração final . . . 133

Referências . . . 135
ANEXO A Produção técnica e acadêmica . . . 149

Dedico esta tese a todos os animais que contribuem para o progresso da ciência.

Agradecimentos

Inicialmente, meu agradecimento mais que especial a meus pais e a minha família pelo apoio e amor incondicional e por estarem sempre tão perto mesmo estando tão longe.

Ao Pedro, por todo o amor, carinho e amizade e pelos sonhos que traçamos juntos. Agradeço pela compreensão nos dois anos de relacionamento à distância, por todas as horas no telefone e pela sua disposição em aprender neurociência junto comigo.

À Lola, por ser minha companhia fiel, por esperar pacientemente que eu acabasse minhas obrigações para que, finalmente a levasse para passear e por diariamente me recordar quanta alegria é estar na presença de quem amamos.

A todos que tive a oportunidade de conhecer nesses dois anos que morei em Natal. Muito obrigada por me deixarem fazer parte de suas vidas e me mostrarem a grandeza da hospitalidade nordestina. Agradeço ao Mauricio, à Lílian e à Camila, por serem minha segunda família aqui. A meus queridos colegas de mestrado e a todos os alunos que estiveram no instituto, pelos abraços, pelo companheirismo e por me recordarem todos os dias que a colaboração é a chave para que alcancemos nossos objetivos. À Célia, à Lúcia, à Bel, ao Firmino e ao Márcio por todas as comidinhas e bate-papos que trouxeram muito mais felicidade a meus dias. À Adriana, ao Aldomar, à Alessandra e ao CEBB por me auxiliarem a enxergar a vida com outros olhos e contribuírem com meu amadurecimento. Sentirei imensas saudades de todos vocês, muito obrigada por tudo.

A meus queridos amigos que moram em outras cidades. Agradeço pela presença e pela amizade, ainda que eu desaparecesse por meses a fio. Um agradecimento especial à Claudia, à Ale, à Gabi, à Ysa, e ao Sônia.

Para finalizar, agradeço a todos os funcionários e professores do instituto, por me auxiliarem ao longo de todo o curso. Agradeço especialmente ao Renan pela paciência infinita, pelo zelo, por toda a ajuda e orientação e por não me deixar desanimar mesmo quando tudo parecia não ter mais jeito.

—

"What is real? How do you define 'real'? If you're talking about what you can feel, what you can smell, what you can taste and see, then 'real' is simply electrical signals interpreted by your brain."- Morpheus (Matrix).

Lista de ilustrações

Figura 1.1 – Esquema representando o potencial de ação da membrana celular ao longo do tempo. Em 1) a membrana está polarizada e em repouso pela ação da bomba de sódio e potássio. Após um estímulo causar uma perturbação que cruza o limiar de excitação em 2) os canais de sódio se abrem, o que causa a entrada desses íons na célula. Aproximando-se de 3), os canais de potássio começam a se abrir e os de sódio fecham. Já em 4), os canais de potássio se fecham e a célula vai lentamente retornando ao potencial de repouso. 5

Figura 1.2 – Representação de um trem de *spikes* e algumas de suas propriedades: Janela ou *bin*, ISI, taxa de disparos e taxa média de disparos. Cada barra vertical marca o instante de tempo em que ocorreu um potencial de ação. 7

Figura 2.1 – Disposição relativa dos neurônios em cada categoria de AN gerada artificialmente. Na figura, temos ANs disjuntas, com pelo menos um neurônio exclusivo e sem neurônios disjuntos, respectivamente. 14

Figura 2.2 – Representação da geração de trens de *spike* a partir da distribuição de Poisson. As barras verticais simbolizam os valores instantâneos da taxa de disparos. Interações são inseridas por modificações pontuais em instantes aleatórios dos trens relativos aos neurônios que participam da AN. À esquerda da figura, exemplificamos uma interação excitatória simulada, em que há um aumento da taxa de disparo e, à direita, uma inibitória, em que há diminuição. 16

Figura 2.3 – *Kernel* do tipo Gaussiano em torno de um ponto de referência (marcado com um "x"). Os círculos marcados na figura mostram que na vizinhança do ponto de referência a função assume um valor limite (em vermelho) que decai à medida que nos distanciamos (em preto). 19

Figura 2.4 – Na figura, as barras contínuas marcam a ocorrência de um *spike* para as células 1, em azul, e 2, em laranja. A barra hachurada em preto marca o instante de tempo de referência, t. As setas ilustram os intervalos de tempo empregados no cálculo da distância SPIKE. 26

Figura 2.5 – Representação esquemática do efeito da resolução na determinação de comunidades. 30

Figura 2.6 – Padrão característico de uma assembleia: os membros apresentam pesos parecidos enquanto que os outros neurônios aparecem com pesos nulos. As Figuras 2.6a e 2.6c apresentam PCs significativas para as ANs disjuntas e exibem um exemplo em que a distribuição dos pesos permite a identificação dos membros. No entanto, as PCs das ANs que compartilham neurônios (Figuras 2.6b e 2.6d) tem uma distribuição não característica e a classificação dos neurônios não pode mais ser determinada. 33

Figura 2.7 – Representação de uma cenário de aplicação do ICA. Três neurônios são gravados por três eletrodos distintos. O sinal obtido por cada eletrodo é uma combinação do sinal de cada uma das células. 34

Figura 2.8 – A figura mostra o resultado para a simulação de três ANs, uma independente e as outras duas compartilhando células. Após a aplicação do PCA ou SVD foi possível observar que o número de autovalores maiores que o limite superior da distribuição de Marčenko-Pastur foi igual a três e que, portanto, o número de assembleias foi corretamente identificado. No entanto, o padrão dos pesos das PCs associadas não permite apontar de maneira imediata a filiação das células para as comunidades que compartilham neurônios, Figuras 2.8a, 2.8c e 2.8e. Para o mesmo conjunto de dados, esse problema é solucionado após a aplicação do ICA e os pesos se distribuem aproximadamente segundo o padrão característico, Figuras 2.8b, 2.8d e 2.8f. 36

Figura 2.9 – Representações de uma rede utilizando grafos. Os nós são representados pelos círculos e as interações entre eles, os links, pelos traços, no caso de interações simultâneas e não dirigidas (Figura 2.9a) e por setas no caso de interações direcionadas (Figura 2.9b). 37

Figura 2.10–Exemplo de dendrograma, gerado a partir da matriz de similaridade descrita na Seção 2.3.4 e utilizando o método de ligação simples. 41

Figura 2.11–Representação esquemática da ligação do tipo simples. A distância entre comunidades é calculada como a menor distância entre membros (Figura 2.11a). Agrupamento distorcido, que pode conter objetos heterogêneos entre si (Figura 2.11b). 42

Figura 2.12–Exemplo de dendrograma, gerado a partir da matriz de similaridade descrita na Seção 2.3.4 e utilizando o método de ligação completa. 43

Figura 2.13–Representação esquemática da ligação completa, em que a distância entre duas comunidades é computada em relação aos objetos mais distantes entre si. Dessa forma, o agrupamento é sensível a *outliers*. 43

Figura 2.14–Representação esquemática da ligação média, em que a distância entre duas comunidades é computada a partir da média das distâncias entre os objetos que as compõe (Marcada por um 'X'). 44

Figura 2.15–Dendrogramas gerados a partir da matriz de similaridade descrita na Seção 2.3.4, utilizando as ligações médias simples (Figura 2.15a)e ponderada (Figura 2.15b). 44

Figura 2.16–Representação esquemática de uma matriz segundo um espaço de estados. A similaridade entre diferentes objetos passa a ser encarada como a probabilidade de transição de um nó para outro. Para facilitar a visualização, setas mais grossas correspondem a probabilidades de transição maiores. Note que a soma das probabilidades de transição é igual a 1. Os nós A,b e C formam uma partição estável ao longo do tempo de Markov, definindo uma comunidade. 48

Figura 2.17–Desenho esquemático do aparato comportamental. O animal distinguia entre aberturas largas e estreitas utilizando suas vibrissas e associava cada cenário à obtenção de uma recompensa em um lado da câmara. Caso a abertura fosse larga, a recompensa líquida seria liberada no lado direito, se estreita, no lado esquerdo. 54

Figura 3.1 – Representação esquemática utilizada para a disposição dos resultados das simulações. Avaliamos o desempenho de cada forma de processamento cobrindo uma grande variedade de cenários, por meio de rodadas extensivas de simulação. As interações entre os neurônios que compõem as AN foram descritas por variações locais da taxa média de disparos, que podem ser visualizadas ao longo dos eixos da Figura. Inibições foram caracterizadas por valores negativos e excitações, por valores positivos. 59

Figura 3.2 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1, linhas 1, 3 e 5) e na classificação dos neurônios (Seção 2.4.2, linhas 2, 4 e 6) quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 30 células e ANs independentes, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 61

Figura 3.3 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1, linhas 1, 3 e 5) e na classificação dos neurônios (Seção 2.4.2, linhas 2,3 e 6) quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 30 células e ANs independentes, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 63

Figura 3.4 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1, linhas 1, 3 e 5) e na classificação dos neurônios (Seção 2.4.2, linhas 2,3 e 6) quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 64

Figura 3.5 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1, linhas 1, 3 e 5) e na classificação dos neurônios (Seção 2.4.2, linhas 2,3 e 6) quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 66

Figura 3.6 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1, linhas 1, 3 e 5) e na classificação dos neurônios (Seção 2.4.2, linhas 2,3 e 6) quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. A atividade de 5 neurônios de uma das ANs foi atrasada um *bin*. 67

Figura 3.7 – Desempenho do ICA associado a diferentes métricas de similaridade quando aplicado a trens artificiais que modelam ANs independentes gerados segundo o modelo de correlação entre os neurônios (MACKE *et al.*, 2009) com uma taxa de disparos basal de 22,37 *spikes/segundo*. Na primeira linha, exibimos o desempenho para o número de assembleias detectadas (Equação 2.37)e, na segunda, para a classificação dos neurônios (Equação 2.4.2). As linhas pontilhadas pretas em cada um dos eixos delimitam os 39 valores distintos de correlação cruzada (extraídos de dados reais) e foram inseridas para salientar que os valores de correlação não eram uniformemente distribuídos. 69

Figura 3.8 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 30 células e ANs com 5 neurônios disjuntos, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. 70

Figura 3.9 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira e na terceira linha, e, na classificação dos neurônios, na segunda e na quarta linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 30 células e ANs com 5 neurônios disjuntos, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. 71

Figura 3.10– Desempenho do ICA associado a diferentes métricas de similaridade quando aplicado a trens artificiais que modelam ANs com 5 neurônios não exclusivos gerados segundo o modelo de correlação entre os neurônios (MACKE *et al.*, 2009) com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Na primeira linha, exibimos o desempenho para o número de assembleias detectadas (Equação 2.37)e, na segunda, para a classificação dos neurônios (Equação 2.4.2). As linhas pontilhadas pretas em cada um dos eixos delimitam os 39 valores distintos de correlação cruzada (extraídos de dados reais) e foram inseridas para salientar que os valores de correlação não são uniformemente distribuídos. 73

Figura 3.11–Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 30 células e ANs sem neurônios disjuntos, sendo a primeira comunidade com 10 neurônios e, a segunda, com 5. Em um total de 10^4 amostras, selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. 74

Figura 3.12–Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira e na terceira linha, e, na classificação dos neurônios, na segunda e na quarta linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 30 células e ANs sem neurônios disjuntos, sendo a primeira comunidade com 10 neurônios e, a segunda, com 5. Em um total de 10^4 amostras, selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. 75

Figura 3.13–Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira e na terceira linha, e na classificação dos neurônios (Seção 2.4.2), na segunda e na quarta, quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 60 células e ANs sem neurônios disjuntos, sendo a primeira comunidade com 25 neurônios e, a segunda, com 10. Em um total de 10^4 amostras, selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. . . 77

Figura 3.14–Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira e na terceira linha, e na classificação dos neurônios (Seção 2.4.2), na segunda e na quarta, quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células e ANs sem neurônios disjuntos, sendo a primeira comunidade com 25 neurônios e, a segunda, com 10. Em um total de 10^4 amostras, selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. 78

Figura 3.15–Desempenho do ICA associado a diferentes métricas de similaridade quando aplicado a trens artificiais que modelam ANs sem neurônios disjuntos gerados segundo o modelo de correlação entre os neurônios (MACKE *et al.*, 2009) com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Na primeira linha, exibimos o desempenho para o número de assembleias detectadas (Equação 2.37) e, na segunda, para a classificação dos neurônios (Equação 2.4.2). As linhas pontilhadas pretas em cada um dos eixos delimitam os 39 valores distintos de correlação cruzada (extraídos de dados reais) e foram inseridas para salientar que os valores de correlação não são uniformemente distribuídos. 79

Figura 3.16–Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios (Seção 2.4.2), na segunda linha, quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 30 células e ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 81

Figura 3.17–Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 30 células e ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. 82

Figura 3.18–Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. 83

Figura 3.19–Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 84

Figura 3.20–Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. A atividade de 5 neurônios de uma das ANs foi atrasada um *bin*. 84

Figura 3.21–Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células com ANs com 5 neurônios em comum. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. 85

Figura 3.22–Representação esquemática do efeito da resolução na determinação de comunidades. 87

Figura 3.23–Dendrogramas resultantes do uso da correlação com a ligação do tipo simples para uma população de 30 células com taxa média de disparos de 22,37 *spikes/segundo* e ANs independentes. Os tipos de interação entre as ANs variam conforme descrito na legenda. 88

Figura 3.24–Dendrogramas resultantes do uso da correlação com a ligação do tipo simples para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Os tipos de interação entre as ANs variam conforme descrito na legenda. 89

Figura 3.25–Dendrogramas resultantes do uso da correlação com diferentes tipos de *Linkage* para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Células Independentes. 90

Figura 3.26–Dendrogramas resultantes do uso da correlação com diferentes tipos de *Linkage* para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Assembleia 1 com interações de excitação e Assembleia 2, de inibição. 90

Figura 3.27–Dendrogramas resultantes do uso de todas as métricas de similaridade com diferentes a ligação simples para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Células Independentes. 92

Figura 3.28–Dendrogramas resultantes do uso de todas as métricas de similaridade com diferentes a ligação simples para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Assembleia 1 com interações de excitação e, assembleia 2, de inibição. 93

Figura 3.29–Dendrogramas resultantes do uso de todas as métricas de similaridade com diferentes a ligação média simples para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Células Independentes. 94

Figura 3.30–Dendrogramas resultantes do uso de todas as métricas de similaridade com diferentes a ligação média simples para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Assembleia 1 com interações de excitação e, assembleia 2, de inibição. . . 95

Figura 3.31–Dendrogramas resultantes do uso da correlação com diferentes tipos de *Linkage* para uma população de artificial de 30 células com taxa média de disparos de 22,37 *spikes/segundo* geradas segundo o modelo de Macke *et al.* (2009) e ANs independentes. Células Independentes. 96

Figura 3.32–Dendrogramas resultantes do uso da correlação com diferentes tipos de *Linkage* para uma população de artificial de 30 células com taxa média de disparos de 22,37 *spikes/segundo* geradas segundo o modelo de Macke *et al.* (2009) e ANs independentes. Assembleias com interações de excitação. 96

Figura 3.33–Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 30 células com ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 98

Figura 3.34–Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 30 células com ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 98

Figura 3.35–Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 60 células com ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 99

Figura 3.36–Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células com ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 100

Figura 3.37–Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células com ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.A atividade de 5 neurônios de uma das ANs foi atrasada um *bin*. 101

Figura 3.38–Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células com ANs com 5 neurônios em comum. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator. 101

Figura 3.39–PCs resultantes da aplicação da correlação associada ao ICA nos registros eletrofisiológicos extracelulares do Rato 1. Os neurônios k, para $k = \{1, ...n_{Neurons}\}$, cujos pesos w_k foram maiores que o limiar (em vermelho) foram considerados como membros da assembleia neural. 104

Figura 3.40–Resultado da aplicação do ICA associado à correlação (Figura 3.40a) e ao ISI (Figura 3.40b) para o Rato 1. Para cada repetição do experimento, segmentos de dados foram selecionados ao redor da discriminação tátil. . 105

Figura 3.41–Histograma centrado no evento de discriminação tátil para as 14 células do córtex somatossensorial primário do Rato 1. Observamos que há modulação similar das células que tanto a métrica de correlação como a métrica ISI associadas ao ICA indicaram pertencer a uma mesma AN. 106

Figura 3.42–Resultado da aplicação da Modularidade associada ao cosseno com um *kernel* exponencial (Figura 3.42a) e a métrica de Billeh associada à Estabilidade de uma partição (Figura 3.42b) para o Rato 1. Para cada repetição do experimento, segmentos de dados foram selecionados ao redor da discriminação tátil. É possível notar que células independentes ou que pertençam simultaneamente às duas comunidades não são verificadas. . . 108

Lista de tabelas

Tabela 3.1 – Caracterização dos dados experimentais analisados. 102
Tabela 3.2 – Correspondência entre o uso das métricas de similaridade correlação e ISI associadas ao ICA para o conjunto de dados dos animais 1,2 e 3. 105
Tabela 3.3 – Correspondência entre o uso das métricas de similaridade correlação e ISI associadas ao ICA para o conjunto de dados dos animais 4,5 e 6. 106
Tabela 3.4 – Áreas que possuem células engajadas na formação de cada uma das ANs. Resultado para a aplicação da Correlação associada ao ICA para os animais 1, 2 e 3. 107
Tabela 3.5 – Áreas que possuem células engajadas na formação de cada uma das ANs. Resultado para a aplicação da Correlação associada ao ICA para os animais 4,5 e 6. 107
Tabela 4.1 – Sumário das simulações realizadas neste trabalho ao combinarmos diferentes métricas de similaridade e métodos de classificação na caracterização de ANs. 128

Lista de Acrônimos e Abreviações

Σ	Matriz de covariância entre um grupo de células
θ	Ângulo de inclinação da matriz em relação à superfície do cérebro.
μ_t	Taxa de disparos por *bin* de um neurônio
C_v	Coeficiente de variação
AN	Assembleia neural
AP	Eixo anteroposterior
DV	Eixo dorsoventral
ICA	Análise de componentes independentes
ISI	Intervalo de tempo entre *spikes*
LFP	Potencial de campo local
ML	Eixo mediolateral
PC	Componente principal
PCA	Análise de componentes principais
PFC	Córtex pré-frontal
PPC	Córtex parietal posterior
S1	Córtex somatossensorial primário
SVD	Decomposição em valores singulares
TMS	Estimulação magnética transcraniana
V1	Córtex visual primário

1 Introdução

O objetivo deste trabalho foi a análise comparativa do desempenho e das limitações de algumas propostas de algoritmos para detecção de assembleias neurais. Para tanto, construimos conjuntos de dados artificiais com características de populações neuronais reais, e analisamos o desempenho desses algoritmos na sua proposta original bem como em variações inéditas das métricas de similaridade e do método de detecção de padrões. O fator determinante no julgamento da qualidade do processamento foi sua capacidade de recuperar as assembleias geradas artificialmente quanto ao número de populações e aos neurônios associados a elas. Além disso, aplicamos as propostas mais promissoras a um conjunto de neurônios de ratos registrados durante a execução de uma tarefa de discriminação tátil, visando uma análise da dinâmica das assembleias neurais dos córtices pré-frontal, parietal posterior, visual primário e somatossensorial primário. Nossas análises destacaram a superioridade de determinados algoritmos na caracterização de assembleias neurais e clarificaram os cenários em que seu uso é mais apropriado. Assim, os resultados apresentados nessa dissertação contribuem para substanciar a escolha da metodologia utilizada na análise dos registros eletrofisiológicos extracelulares de uma população de células.

Este capítulo apresenta a motivação desse trabalho e inicia-se com um breve histórico sobre a teoria de assembleias neurais. Na sequência, descreveremos o cenário atual de processamento de trens de *spikes* na neurociência salientando a crescente importância de algoritmos capazes de identificar e monitorar células responsáveis pela codificação de informação. O capítulo encerra-se com o detalhamento dos objetivos da dissertação.

1.1 Assembleias Neurais

Camillo Golgi foi um cientista italiano que contribuiu de maneira decisiva para o início da neurociência moderna. Ele criou um método revolucionário capaz de marcar a membrana celular, a chamada "reação negra", uma solução de dicromato de potássio e nitrato de prata que se acumulava no corpo celular, nos dendritos e nos axônios, delimitando claramente a morfologia dos neurônios. Isso permitiu, de maneira pioneira, que as complexas ramificações do tecido nervoso fossem visualizadas em detalhes. Sua técnica, no entanto, não permitia a observação de uma separação clara entre células, o que levou Golgi a interpretar seus resultados afirmando que o tecido nervoso formava redes reticulares, ou seja, que o cérebro é formado por uma rede única e contínua interligada através de ramificações do axônio

(BENTIVOGLIO, 1998a; SABATTINI, 2003).

Santiago Ramón y Cajal, um cientista espanhol, aprimorou as técnicas de Golgi e, dentre outras contribuições, fez uma série de ilustrações, estudando a estrutura de vários tipos de células nervosas e suas interconexões em diversas áreas do sistema nervoso central para diferentes espécies. No entanto, diferentemente de Golgi, ele era um defensor da denominada teoria neuronal. A teoria argumentava que os neurônios eram independentes e constituíam a unidade estrutural e funcional do sistema nervoso, com capacidade de se comunicarem entre si (BENTIVOGLIO, 1998b; SABATTINI, 2003).

Apesar de dividirem o prêmio Nobel em Fisiologia ou Medicina em 1906, os dois cientistas não chegaram a um consenso. Posteriormente, experimentos com microscopia eletrônica vieram a comprovar que, de fato, Ramón y Cajal estava correto e que os neurônios são células independentes. Além disso, a "lacuna"entre células vizinhas, ou sinapse, por onde o sinal elétrico é transmitido de uma célula para outra pôde ser analisada morfologicamente, com a visualização das vesículas de neurotransmissores químicos (SABATTINI, 2003).

Com a força da teoria neuronal, iniciou-se uma frente de pesquisa em neurociência que buscava a compreensão do cérebro por meio do entendimento de cada tipo de neurônio individualmente, de acordo com sua localização e propriedades fisiológicas. Neste cenário, pouco se especulou sobre a maneira como essas células trabalhariam de maneira integrada ("*binding problem*", na literatura em inglês, ver Seção 1.3) e, principalmente, nos casos com o animal em alerta ao invés dos tradicionais experimentos sob anestesia.

Décadas depois, Hebb (1949) retomou algumas das ideias de Ramón y Cajal, mas propôs que as sinapses constituíam a base do aprendizado associativo e que modificações a nível celular acarretariam em alterações metabólicas, resultando em mudanças nos padrões de atividade de um conjunto de células neurais, ou assembleias neurais (ANs). Essas populações se associariam por conexões sinápticas flexíveis formando "*microcirtuitos*"de processamento de informação distribuída. Assim, interações se dariam de maneira funcional, e não exclusivamente estrutural (SAKURAI, 1998). Com isso, Hebb estimulou a realização de experimentos com múltiplos neurônios cuja atividade elétrica é registrada simultaneamente em contrapartida àqueles em que um único neurônio é gravado em múltiplas sessões (QUIANQUIROGA; PANZERI, 2009).

Mais precisamente, o modelo sináptico hebbiano consiste num mecanismo interativo com dependência temporal e de atuação local que aumenta sua eficiência em função da correlação entre as atividades pré e pós sinápticas (HAYKIN, 1999), ou seja, a sinapse é fortalecida, caso a correlação entre o sinal pré e pós sináptico seja positiva, ou enfraquecida, caso contrário.

A descrição de Hebb levou à formulação da hipótese das assembleias neurais, que define uma população, e não um único neurônio, como a unidade fundamental no processamento de sinais cerebrais (NICOLELIS *et al.*, 1997; HARRIS, 2005; BUZSáKI, 2010). A teoria argumenta que neurônios individuais não poderiam descrever fenômenos complexos do cérebro, como, por exemplo, redundância e aprendizado, já que o disparo de uma célula depende da atividade das milhares de células às quais ela está conectada e, portanto, não é determinístico (GERSTEIN, 1989; STERIADE *et al.*, 1993; SAKURAI, 1998).

Em resumo, os principais pontos da teoria hebbiana são a redundância de codificação de informação [1], o processamento distribuído, a formação e a reconfiguração dinâmica de ANs [2] e, finalmente, a ativação simultânea [3] (SAKURAI, 1998). Entretanto, aspectos envolvendo a dinâmica temporal de formação e, principalmente, de dissolução de assembleias, que explicassem o fato de o cérebro não convergir a estados inconsistentes, como uma sincronização global não foram teorizados por Hebb (BUZSáKI, 2006). Nesse sentido, a proposta hebbiana foi ligeiramente modificada e, o conceito de AN foi expandido para englobar não apenas células com conexões sinápticas entre si, mas todas as células que apresentem interações maiores que o acaso. Dessa maneira, o conceito de ANs torna-se mais geral e abrange até mesmo neurônios distantes e de diferentes áreas que porventura estejam associados na codificação da informação, seja por meio de circuitos complexos, seja pelo recebimento simultâneo de estímulos em comum. Além disso, essa definição em larga escala incorpora também a noção de que células relacionadas e não síncronas, ou seja, com atrasos temporais entre si, podem pertencer à mesma AN (VARELA *et al.*, 2001).

A hipótese da existência de ANs é amplamente discutida na comunidade científica, já que corrobora com uma série de fenômenos fisiológicos e psicológicos (GERSTEIN, 1989), como, por exemplo, a plasticidade funcional, a cognição, o planejamento, a atenção e a recuperação de memórias (SAKURAI, 1998; FRIES, 2005; BUEHLMANN; DECO, 2008; CISEK; KALASKA, 2010; ENGEL *et al.*, 2013; QIN; YU, 2013). Além disso, a atividade de populações neuronais já foi empregada com sucesso em aplicações em tempo real, em especial em experimentos que envolvem a decodificação da intenção motora de um animal e posterior transmissão para interfaces cérebro-máquina, como, por exemplo, Chapin *et al.* (1999), Wessberg *et al.* (2000), Serruya *et al.* (2002), Carmena *et al.* (2003), Hochberg *et al.* (2006), Hochberg *et al.* (2012), Aggarwal *et al.* (2013), Collinger *et al.* (2013), Shokur *et al.* (2013), Ifft *et al.* (2013), e até mesmo entre cérebros de animais distintos (PAIS-VIEIRA *et*

[1] A redundância segundo a teoria hebbiana sugere que um único neurônio pode participar de múltiplas ANs.

[2] Segundo a teoria de Hebb, a formação e a reconfiguração dinâmica de ANs dependerão da realimentação que manterá ou não a interação sináptica entre as células.

[3] A ativação simultânea prevê que, uma vez que um subconjunto mínimo da AN esteja atuante, isso acarretará no acionamento de todas as células da assembleia.

al., 2013b).

No entanto, a confirmação experimental dessa hipótese ainda é um dos principais desafios da neurociência (BUZSáKI, 2010). Uma das limitações nesse sentido, a aquisição simultânea do sinal elétrico de várias células, apresentou progressos significativos nos últimos anos (NICOLELIS *et al.*, 1997; BUZSáKI, 2010; SCHWARZ *et al.*, 2014), mas o crescente número de neurônios registrados enfatiza a necessidade de novos métodos matemáticos e estatísticos que possam lidar com uma grande quantidade de dados, processando a atividade simultânea de muitos neurônios e detectando ANs. A pesquisa de novos métodos é relevante tanto num cenário em que buscamos implementar e aprimorar a resolução de interfaces cérebro-máquina em tempo real quanto em análises *offline*, já que a maneira com que diferentes áreas cerebrais interagem no processamento da informação ainda está sendo estabelecida.

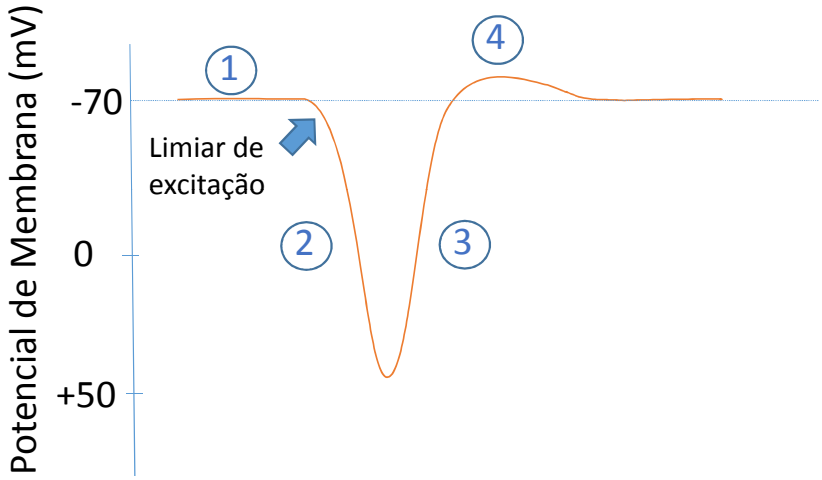
Uma vez que a AN seja encarada como a unidade funcional de informação do sistema nervoso, sua dinâmica será definida pela maneira com que a informação for codificada pelo cérebro. Assim, o entendimento das teorias do código neural faz parte do estudo de assembleias e será explorado na Seção 1.3, precedida por uma seção introdutória sobre sinais neurais e terminologias utilizadas ao longo desse trabalho.

1.2 Trens de *spike*: Propriedades e terminologias

Nesta seção descreveremos brevemente os sinais neurais envolvidos nesse trabalho, assim como definiremos as terminologias utilizadas nas próximas seções.

- Potencial de ação ou disparo (*spike*, na literatura em inglês): O potencial da membrana depende da ação de proteínas presentes na membrana celular, que regulam o fluxo de íons entre os meios extra e intra celular. Essas proteínas podem ser classificadas em duas categorias: os canais iônicos, e os transportadores ou bombas. Os canais iônicos podem ser encarados como pequenos poros de permeabilidade seletiva, distribuídos de maneira heterogênea ao longo da superfície celular, que respondem com uma dinâmica rápida e muito específica a estímulos químicos e físicos, controlando a entrada e saída de íons. Existem vários canais, que se distinguem quanto aos íons a que são seletivos, quanto a sua dinâmica e quanto aos fatores que controlam sua permeabilidade. Os canais com maior impacto na morfologia do *spike* são os de sódio e potássio. As bombas, por sua vez, tem uma dinâmica lenta. Elas são importantes na manutenção dos gradientes de concentração entre os meios extra e intracelular. Uma das mais importantes é a bomba de sódio e potássio, um mecanismo de transporte ativo, aumentando a concentração

Figura 1.1 – Esquema representando o potencial de ação da membrana celular ao longo do tempo. Em 1) a membrana está polarizada e em repouso pela ação da bomba de sódio e potássio. Após um estímulo causar uma perturbação que cruza o limiar de excitação em 2) os canais de sódio se abrem, o que causa a entrada desses íons na célula. Aproximando-se de 3), os canais de potássio começam a se abrir e os de sódio fecham. Já em 4), os canais de potássio se fecham e a célula vai lentamente retornando ao potencial de repouso.



Fonte: Adaptada de Cofer (2002)

dos íons sódio (Na^+) no meio extracelular e potássio (K^+), no intracelular. A atuação da bomba de sódio e potássio em conjunto com os canais seletivos a esses íons faz com que as células permaneçam polarizadas, mesmo em situações de repouso, num potencial de aproximadamente -70 mV para os neurônios (Figura 1.1, fase 1). Os potenciais de ação são rápidas mudanças no potencial de membrana que, a medida que se propagam, transmitem informação de uma célula para outra, mesmo a grandes distâncias. Quando a excitação da membrana cruza um limiar, há uma rápida despolarização, que corresponde à abertura dos canais de sódio e o influxo desses íons para o interior da célula (Figura 1.1, fase 2). Essa modificação no potencial da membrana aciona a abertura dos canais de potássio e o fechamento dos canais de sódio. A saída dos íons de potássio leva a membrana à repolarização (Figura 1.1, fase 3). Por possuírem uma dinâmica mais lenta, os canais de potássio levam a membrana a uma hiperpolarização (Figura 1.1, fase 4) e a célula lentamente retorna ao potencial de repouso pela ação da bomba de sódio e potássio. Sua escala temporal é da ordem de 1 ms (KANDEL, 2012).

- *Spike sorting*: O *spike sorting*, como é chamado na literatura em inglês, consiste numa etapa do processamento do sinal que busca distinguir a atividade de células distintas, mas também separar o sinal do neurônio do ruído.

- Trem de *spike*: Vetor que armazena as ocorrências de potenciais de ação de um neurônio ao longo do tempo. Pode ser representado graficamente por uma série de barras verticais, como ilustrado na Figura 1.2.

- Janela ou *bin*: Intervalo de tempo utilizado para definir a escala temporal da análise (Figura 1.2).

- Intervalo entre disparos (ISI) (*Inter-Spike Interval*, na literatura em inglês): Corresponde à diferença de tempo entre a ocorrência de dois *spikes* consecutivos (Figura 1.2).

- Taxa de disparos: Número de ocorrências de potenciais de ação de um neurônio em uma janela (Figura 1.2).

- Taxa média de disparos: Valor médio da taxa de disparos em um segmento de dados (Figura 1.2).

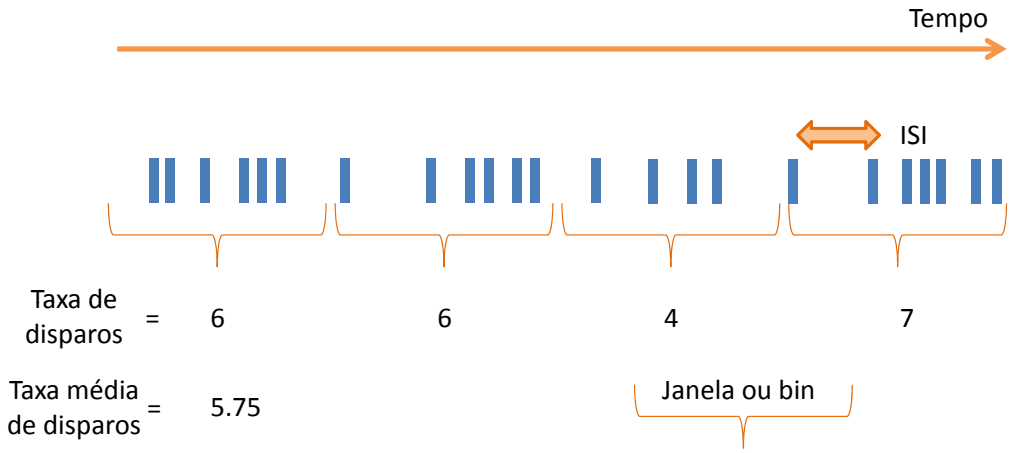
1.3 O código neural

Uma das maneiras de buscar a compreensão de como a informação é codificada pelo cérebro é a estudando a resposta neural em relação a um estímulo de referência. Registros de potenciais de ação levaram à constatação de que seu formato de onda não sofre modificações significativas com a intensidade do estímulo recebido, ou seja, acima de um determinado limiar, *threshold* na literatura em inglês, e representado na Figura 1.1, a resposta em amplitude do potencial de membrana da célula é praticamente inalterada. Essa baixa variabilidade implica em um menor poder de transporte de informação pela forma de onda e códigos baseados nesse paradigma foram preteridos.

No entanto, foi possível observar uma correlação entre variações do estímulo com a frequência de disparos. Sob a ótica de ideias localizacionistas e da teoria neuroral em que os neurônios seriam unidades básicas funcionais de processamento de informação, surgiram as primeiras hipóteses afirmando que variações na atividade da célula seriam um mecanismo de codificação do estímulo de maneira complexa e bem específica. Vários experimentos foram feitos no sentido de descobrir a qual estímulo ou conceito uma célula responderia, como por exemplo as famosas células da vovó (*grandmother cells*, na literatura em inglês), que estariam ativas quando um estímulo associado à avó fosse apresentado (GROSS, 2002).

No entanto, conforme mencionado na Seção 1.1, esse tipo de processamento tem um problema, o paradoxo conhecido como "*binding problem*", pois, afinal, se cada área cortical é especializada e separada espacialmente de outras, de que maneira toda essa informação

Figura 1.2 – Representação de um trem de *spikes* e algumas de suas propriedades: Janela ou *bin*, ISI, taxa de disparos e taxa média de disparos. Cada barra vertical marca o instante de tempo em que ocorreu um potencial de ação.



Fonte: Produzida pela autora

seria agrupada para constituir nossas percepções e nossa consciência? Do ponto de vista de codificação de informação, um modelo em que cada célula codifica um estímulo específico é ineficiente, pois os conceitos apresentam grandes semelhanças entre si. Somado a isso, existem as constatações de que esse paradigma não explica alguns fenômenos, como a plasticidade cerebral, e que ao considerarmos todos os estímulos que um indivíduo recebe ao longo da vida, o número de neurônios presentes no cérebro não seria suficiente para processá-los (SAKURAI, 1998).

Assim, a hipótese de um código cerebral distribuído ganhou força, pois as reconfigurações morfológicas e múltiplas combinações entre células trazem uma possibilidade praticamente ilimitada de codificação da informação. No entanto, o processamento do código neural tendo por base as variações na taxa de disparos permaneceu já que corrobora com uma série de experimentos. Cabe ressaltar que há uma ligeira diferença em relação à teoria original de Hebb, já que, nesse caso, a interação entre células é assumida numa escala temporal maior

(PIPA, 2006). Isso seria vantajoso uma vez que empregar um número maior de *spikes* traz mais robustez contra as variações intrínsecas de um sistema estocástico. Apesar de ser possível a definição de taxas médias de disparo em *bins* menores, isso pode trazer problemas na estimativa da informação, já que a estacionaridade não pode ser garantida (PAIVA *et al.*, 2010) e o trem tenderá a ser muito esparso, causando uma distribuição muito assimétrica (PEYRACHE *et al.*, 2010).

Contudo, observou-se experimentalmente que vários comportamentos dos animais ocorrem numa escala temporal menor que a necessária para a obtenção de um número significativo de *spikes* para uma codificação baseada em taxas. Logo, essa abordagem só é fidedigna em cenários em que a resposta comportamental é lenta (RIEKE *et al.*, 1997). Além disso, constatou-se que estímulos distintos podem estar associados a taxas instantâneas de disparo de mesma magnitude, o que não permitiria a distinção de assembleias distintas (esse fenômeno é conhecido como *superposition catastraphe problem*, na literatura em inglês). Para superar esse problema, foi proposto que as ANs seriam separáveis em análises que consideram diferentes escalas temporais (MALSBURG; SCHNEIDER, 1986).

Porém, nota-se que um código baseado na taxa média de disparos negligencia a estrutura temporal interna do trem de *spike*, e, possivelmente, descarta informação, o que induziu o surgimento de alternativas baseadas no tempo preciso de ocorrência de um potencial de ação. Algumas propostas nesse sentido são: o código baseado na latência de uma célula, ou seja, pelo tempo entre a ocorrência de um estímulo e o início de sua resposta eletrofisiológica; o código baseado no intervalo entre *spikes* consecutivos e, finalmente, um código que combina diferenças de fase entre as ocorrências de potencias de ação individuais e o potencial elétrico produzido por um grande numero de células adjacentes, o denominado potencial de campo local (LFP). Independentemente das vantagens e desvantagens de cada abordagem, o mais provável é que a informação no cérebro seja multiplexada em diferentes escalas temporais (PANZERI *et al.*, 2010), ou seja, que informações complementares estejam codificadas em escalas temporais distintas.

Apesar do código neural ainda ser desconhecido, foram apresentadas na literatura diversas propostas que quantificam a semelhança entre os disparos das células, cada uma baseada em características específicas (Seção 2.2). No entanto, não há consenso sobre qual abordagem produz os melhores resultados. Neste trabalho, selecionamos propostas com diferentes escalas temporais, tanto baseadas na taxa de disparos quanto na estrutura temporal do trem, e fizemos uma análise comparativa no impacto que elas podem exercer no processamento de ANs.

1.4 Assembleias neurais em trens de *spike*

Em geral, as metodologias descritas na literatura para caracterização espaço-temporal de ANs e seus respectivos membros seguem um processamento estruturado, com passos bem definidos (BILLEH *et al.*, 2014). Este trabalho emprega o mesmo conjunto de etapas, descritas a seguir:

- Geração de um conjunto de dados de teste: A primeira etapa geralmente consiste na geração de um conjunto de dados com interações controladas e biologicamente inspiradas. Isso é necessário pois permite a comparação do resultado obtido após o processamento com as populações pré-estabelecidas. A maneira mais empregada consiste em pequenas modificações de trens gerados segundo a modelagem de Poisson (Seção 2.1.2), como os trabalhos de Laubach *et al.* (1999), Feldt *et al.* (2009), Santos *et al.* (2011), Gansel e Singer (2012), Santos *et al.* (2013), Picado-Muino *et al.* (2013). Como alternativas podemos citar a criação de redes artificiais usando modelos simplificados da dinâmica do potencial da membrana, como o integra e dispara (BILLEH *et al.*, 2014) e o de Hindermash-Rose (KREUZ *et al.*, 2007; KREUZ *et al.*, 2009); ou ainda, o empregado no conjunto de dados de Fellous *et al.* (2004), usado também por Humphries (2011), que adotam dados obtidos pela estimulação de células *in vitro*, cuja estrutura anatômica é conhecida. Após a gravação, pequenas perturbações temporais aleatórias são inseridas na ocorrência dos *spikes*.

- Cálculo de uma métrica de similaridade: Essa etapa consiste num pré-processamento dos dados, segundo um padrão de interações entre os neurônios, ou seja, algum aspecto específico da atividade é selecionado para definir a formação ou supressão de associações entre as células ao longo do tempo. Alguns exemplos são a covariância da taxa média de disparos (QUIANQUIROGA; PANZERI, 2009; PANZERI *et al.*, 2010; SANTOS *et al.*, 2011; SANTOS *et al.*, 2013), a coincidência temporal de *spikes* (HARRIS, 2005; QUIANQUIROGA; PANZERI, 2009; PANZERI *et al.*, 2010) e a informação mútua (QUIANQUIROGA; PANZERI, 2009; PANZERI *et al.*, 2010; LONG; CARMENA, 2011).

- Aplicação de um método de classificação: Nesta etapa, aplica-se um algoritmo que visa a identificação de possíveis agrupamentos neurais. Ele recebe como entrada a matriz formada pelas interações pareadas, calculada pela métrica de similaridade. Vários algoritmos empregados são bastante populares na área de aprendizado de máquina. Os mais interessantes para essas aplicações são não supervisionados, uma vez que os dados podem não ser rotulados, e, se possível, não paramétricos, o que diminuiria a influência

do experimentador no resultado. Um exemplo é a adaptação da análise de componentes principais para este problema (Seção 2.3.1).

- Avaliação do desempenho: Para finalizar, o resultado obtido é comparado com o conjunto de dados simulado segundo algum critério, como por exemplo o número de células corretamente classificadas (BILLEH *et al.*, 2014) ou o número de populações identificadas (KREUZ *et al.*, 2007).

1.5 Objetivos da dissertação

Neste projeto de mestrado, abordamos o problema de detecção e identificação dos neurônios que compõe uma AN. Os principais objetivos foram:

- Analisar comparativamente alguns algoritmos já apresentados na literatura, num conjunto de cenários pré-definidos, tal que todos trabalhassem com os mesmos dados de entrada e possuíssem a mesma estrutura de processamento explicada na Seção 1.4.

- Verificar o impacto no desempenho caso métricas de similaridade distintas das originalmente propostas fossem usadas com os métodos de classificação.

- Explorar as limitações inerentes a cada etapa do processamento, a fim de orientarmos experimentadores sobre como empregar uma análise mais eficiente tendo em vista as características dos dados coletados e a aplicação à que esses dados se dirigem. Um entendimento desses pontos é relevante pois mesmo algoritmos considerados mais simples podem exibir resultados promissores no caso de simulações mais realistas (DEOLINDO *et al.*, 2014), o que pode ter um impacto considerável no tempo de processamento de análises de grandes quantidades de dados.

- Considerando-se que a maioria dos métodos existentes pressupõe interações entre neurônios puramente excitatórias e sem atrasos de fase, o projeto propôs estender as análises também para cenários com inibição e com atrasos, dada a sua relevância. Abordamos também os casos de ANs disjuntas (total ou parcialmente) já que o estudo de ANs que compartilham neurônios é pouco explorado e há evidências de que algumas células participam simultaneamente de múltiplas comunidades.

- Por meio de registros eletrofisiológicos de trens de *spike*, analisar a formação dinâmica de ANs em diferentes áreas corticais do cérebro de ratos executando uma tarefa de discriminação tátil. Utilizamos as conclusões obtidas a partir de dados simulados para fundamentar a escolha do método de processamento.

Esta dissertação está organizada da seguinte forma: na Seção 2.1 fazemos uma descrição dos modelos matemáticos empregados na geração de dados artificiais, e retratamos a obtenção de parâmetros que conferem plausibilidade biológica. As métricas de similaridade foram descritas na Seção 2.2 e os métodos de classificação na Seção 2.3. Uma descrição das métricas de avaliação de desempenho segue na Seção 2.4. Na Seção 2.5 descrevemos a tarefa comportamental e os dados eletrofisiológicos que foram analisados. Os resultados foram apresentados no Capítulo 3 e discutidos no Capítulo 4. O Capítulo 5 encerrra o trabalho com as conclusões.

2 Métodos

2.1 Geração de trens de *spike* artificiais

Nesta seção abordaremos a metodologia utilizada na confecção dos dados artificiais empregados em nossas análises. Começaremos explicando a relação entre esses trens e situações reais. A seguir, apresentaremos os métodos selecionados para a geração de trens de *spike* artificiais. Idealmente, os modelos computacionais utilizados devem ser capazes de simular o comportamento esperado com o máximo nível de detalhes possível enquanto mantém um baixo custo computacional. Portanto, selecionamos dois métodos com diferentes níveis de complexidade a fim de julgar o quanto simplificações matemáticas do comportamento de populações pode enviesar nossas conclusões. Ambos são biologicamente plausíveis, sendo o primeiro baseado em trens independentes, e o segundo baseado em variáveis com interações de segunda ordem entre si.

2.1.1 Seleção de características de dados e populações reais.

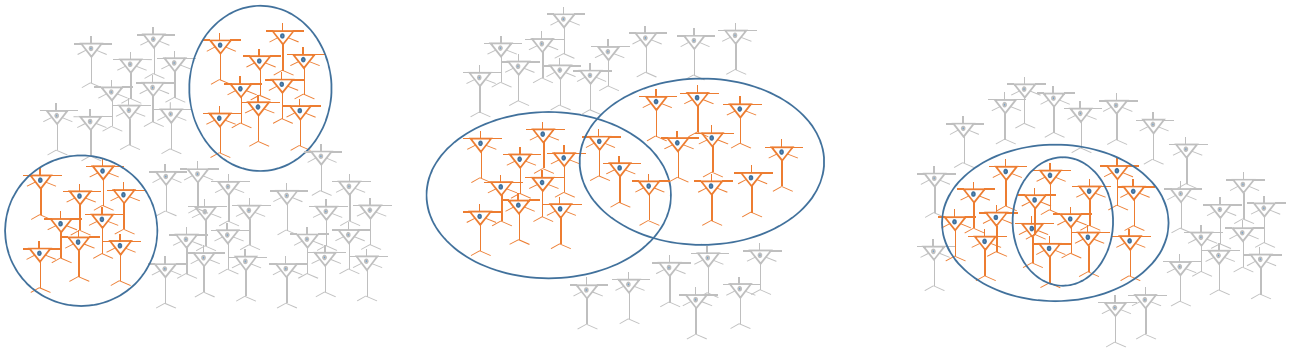
Conforme já brevemente mencionado na Seção 1.2, quando o potencial de membrana cruza um limiar de excitação, ocorre a abertura dos canais de sódio, que leva à ocorrência de um potencial de ação. Dessa maneira, os neurônios eram majoritariamente enxergados como unidades do tipo integra-e-dispara (BRUNEL; ROSSUM, 2007), ou seja, estímulos externos oriundos de diferentes fontes seriam integrados até que o limiar de excitação fosse alcançado, o que levaria à ocorrência de um potencial de ação.

Posteriormente, confirmou-se que o neurônio é uma estrutura mais complexa e que a geração de *spikes* advém da ação combinada de diversas interações não lineares entre canais iônicos, diferenças de potencial e correntes (HODGKIN; HUXLEY, 1990). Verificou-se também que os neurônios podem responder de maneira distinta a um mesmo estímulo (IZHIKEVICH, 2007).

Assim, surgiram implementações computacionais mais completas, que modelam neurônios de diversas áreas, mantendo suas propriedades anatômicas e que incluem uma grande variedade de canais iônicos (MARENCO *et al.*, 2015). Além disso, modelos mais simplificados como o integra-e-dispara e o de Izhikevich (2007) ainda são bastante utilizados em simulações de grandes populações neuronais.

No entanto, esses modelos contém parâmetros variáveis que, ao serem ajustados, mo-

Figura 2.1 – Disposição relativa dos neurônios em cada categoria de AN gerada artificialmente. Na figura, temos ANs disjuntas, com pelo menos um neurônio exclusivo e sem neurônios disjuntos, respectivamente.



Fonte: Produzida pela autora

dificam o padrão de disparos de uma população gerada artificialmente. Assim, a simulação de populações segundo esse paradigma requer um conhecimento *a priori* das áreas que desejamos simular. Ademais, conforme mencionado na Seção 1.3, acredita-se que a morfologia do potencial de ação carrega pouca informação sobre os estímulos. Como o objetivo deste trabalho é testar métodos de detecção de ANs para registros eletrofisiológicos extracelulares de trens de *spike* de qualquer região, optamos pelo uso de modelos que simulam a ocorrência de *spikes* como eventos binários segundo modelos estatísticos.

Para produzir dados artificiais com propriedades de dados biológicos reais, utilizamos como referência os disparos neuronais de um animal em vigília. Os parâmetros escolhidos para representar essa população foram suas características estatísticas mais simples, de primeira e segunda ordem: a taxa média de disparos por *bin* (μ_t) e a matriz de covariância entre o grupo de células (Σ).

Outro fator decisivo para os dados simulados é a especificação das ANs, definindo os neurônios que pertencem a cada comunidade artificialmente gerada. Conforme abordado na Seção 1.3, a teoria mais aceita na comunidade científica defende que as células neurais são capazes de multiplexar no tempo o processamento de informação. Dessa maneira, há evidências de que comunidades reais não sejam disjuntas e de que algumas células participem simultaneamente de múltiplas ANs, como prevê a teoria de Hebb (SAKURAI, 1998). Portanto, seguindo a proposta de Santos *et al.* (2013), nossas assembleias artificiais foram classificadas em três categorias: disjuntas, com pelo menos um neurônio exclusivo e sem neurônios disjuntos, conforme pode ser verificado na Figura 2.1

Além disso, executamos simulações em que foram definidos atrasos relativos entre os membros da assembleia, o que caracteriza também diferenças de fase não nulas.

É importante ressaltar que grande parte dos algoritmos para detecção de ANs aborda apenas o contexto de ANs disjuntas, sem atraso de fase, e puramente excitatórias. Neste projeto, analisamos também os casos de ANs disjuntas (total ou parcialmente), com interações inibitórias, e com atraso de fase.

Na próxima seção descreveremos com mais detalhes os modelos matemáticos empregados no modelamento dos disparos de populações neuronais.

2.1.2 Geração de trens de *spike* a partir da distribuição de Poisson

Experimentos neurofisiológicos mostram que caso a resposta de um neurônio seja registrada quando ele for submetido a um mesmo estímulo em diversas tentativas, muito dificilmente a resposta obtida em cada tentativa será exatamente igual entre elas. Dessa maneira, o modelamento de trens de *spike* começa necessariamente pela consideração de um processo probabilístico.

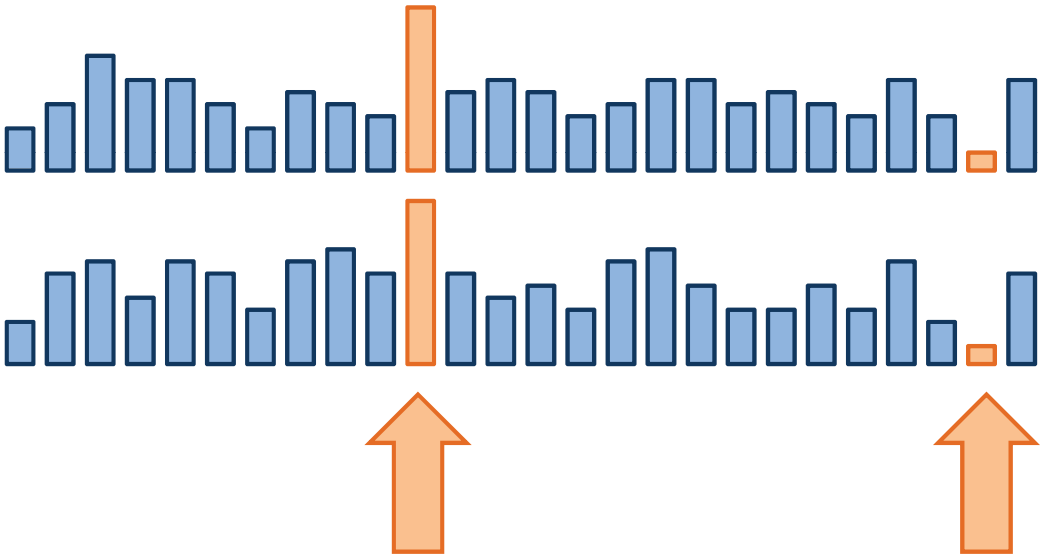
De maneira geral, na teoria de processos estocásticos, a probabilidade de um evento ocorrer num dado instante de tempo depende de todo o histórico de ocorrência de eventos precedentes. Se, para um caso específico, considerarmos que a ocorrência desse evento não tem qualquer relação com os precedentes, teremos um evento estatisticamente independente, num processo de Poisson, que consiste numa generalização de processos de Bernoulli. O processo pode ser do tipo homogêneo, no qual sua média é constante ao longo do tempo, e não homogêneo, em que o valor da média tem dependência temporal. O modelo poissonico consiste numa boa aproximação do disparo de células neurais (DAYAN; ABBOTT, 2001), e o processo homogêneo é bastante empregado devido à simplificação de independência temporal (LAUBACH *et al.*, 1999; HARRIS, 2005; AMARASINGHAM *et al.*, 2006).

As relações de dependência entre os neurônios selecionados para compor cada assembleia foram inseridas através de alterações pontuais na taxa média de disparos em *bins* selecionados aleatoriamente. Essa variação foi positiva quando simulando interações excitatórias e negativa para inibições (Figura 2.2).

2.1.3 Geração de trens de *spike* a partir da correlação entre neurônios

Apesar de ser uma boa aproximação, o processo poissonico não consegue modelar toda a variabilidade de padrões observada em disparos neurais. É possível que, por exemplo, existam correlações entre os disparos de algumas células, modulados por algum fator, como ocorre com células do córtex visual primário (V1) ao serem submetidas a estímulos visuais. Esses desvios podem ser verificados por meio de algumas medidas, como o fator de Fano e o

Figura 2.2 – Representação da geração de trens de *spike* a partir da distribuição de Poisson. As barras verticais simbolizam os valores instantâneos da taxa de disparos. Interações são inseridas por modificações pontuais em instantes aleatórios dos trens relativos aos neurônios que participam da AN. À esquerda da figura, exemplificamos uma interação excitatória simulada, em que há um aumento da taxa de disparo e, à direita, uma inibitória, em que há diminuição.



Fonte: Produzida pela autora

coeficiente de variação (C_v), duas medidas de dispersão, e pela distribuição dos intervalos de tempo entre *spikes* (ISI) (GRU; GRüN, 2009).

A fim de simular uma atividade com padrões temporais mais ricos, incluindo também interações de ordem superior, fizemos simulações utilizando um modelo gaussiano com multivariáveis latentes seguindo a metodologia proposta por Macke *et al.* (2009). Isso é relevante porque um modelo com correlações corresponde melhor a dados reais do que um com variáveis independentes (MACKE *et al.*, 2009), e dados não fidedignos podem levar a falsas conclusões quando aplicamos algoritmos de detecção de ANs (GRU; GRüN, 2009; DEOLINDO *et al.*, 2014).

Esse modelo é interessante por possuir poucos parâmetros e exibir valores de entropia muito próximos aos máximos teóricos numa grande variedade de cenários. Isso sugere que a taxa média de disparos (μ_t) e a matriz de correlação (Σ) sejam suficientes para caracterizar as propriedades estatísticas de uma população de neurônios (MACKE *et al.*, 2009).

Para esse caso, assumimos que a conectividade funcional era estabelecida entre as diferentes unidades por meio da adição de dependências estatísticas (BUZSáKI, 2010). As ANs foram especificadas por meio da covariância entre neurônios, onde valores não nulos

definiam a existência de uma interação.

É importante ressaltar que apesar de capturarmos valores biologicamente inspirados para definir a correlação entre os neurônios, nem todas as matrizes de covariância são factíveis e estão condicionados à existência de uma solução para o modelo gaussiano discretizado, conforme explicado por Macke *et al.* (2009).

Na próxima seção, abordamos as diferentes métricas de similaridade utilizadas neste projeto.

2.2 Métricas de similaridade entre trens de *spike*

Nesta seção fizemos uma revisão dos diferentes métodos usados na quantificação da similaridade entre pares de trens de *spike*, e que foram avaliados neste trabalho. Uma análise de cada proposta é relevante dado que a natureza do código neural ainda é desconhecida e cada métrica se baseia em características específicas dos trens para definir a formação ou supressão de ANs ao longo do tempo.

Além disso, as ANs são identificadas em relação à métrica escolhida, e, portanto, os métodos de classificação (Seção 2.3) podem apresentar resultados diferentes à medida que se modifica a maneira de mensurar a interação entre células. Alguns trabalhos avaliam esse impacto (KREUZ *et al.*, 2007; PAIVA *et al.*, 2010; HUMPHRIES, 2011; BILLEH *et al.*, 2014), porém, até onde pudemos verificar, uma comparação entre diferentes associações de métricas e métodos de classificação ainda não foi explorada.

As métricas utilizadas nesse trabalho foram selecionadas devido a seu uso em algoritmos já propostos na literatura de detecção e monitoramento de ANs além de extraírem diferentes características dos trens de disparos para caracterizar uma interação entre células.

2.2.1 Métricas de similaridade baseadas na taxa de disparos

Conforme mencionado na Seção 1.3, as primeiras hipóteses de codificação de informação pelo sistema nervoso estavam associadas à modulação da taxa de disparos. Dessa maneira, uma série de medidas de similaridade propostas levam em consideração que os disparos são separados em pequenas janelas de tempo de duração constante, ou *bins*, a fim de que as variações instantâneas dessa grandeza sejam suavizadas. É importante ressaltar que ao alterarmos o tamanho da janela, naturalmente, há uma modificação na taxa de disparos $\mu_t(t)$ (KRUSKAL *et al.*, 2007). Os métodos baseados neste paradigma e analisados neste trabalho foram descritos ao longo desta seção.

2.2.1.1 Correlação

A correlação $\rho_{A,B}$ mensura o quanto duas variáveis aleatórias A e B se modificam simultaneamente e é calculada pela Equação 2.1, em que $E\{X\}$ é o operador valor esperado. É normalizada, ou seja, está no intervalo [-1,1], e o sinal indica se a observação do aumento de uma grandeza relaciona-se com o aumento (sinal positivo) ou diminuição (sinal negativo) da outra. No contexto de detecção de ANs, calcula-se a matriz de correlação a partir da taxa de disparos normalizada (*zscore*), conforme descrito por Laubach *et al.* (1999), Peyrache *et al.* (2010), Santos *et al.* (2011), Santos *et al.* (2013), quantificando as variações concomitantes ao longo do tempo em μ_t de duas células.

$$\rho_{A,B} = \frac{E(AB) - E(A)E(B)}{\sqrt{E(A^2) - E^2(A)}\sqrt{E(B^2) - E^2(B)}} \tag{2.1}$$

2.2.1.2 Hamming

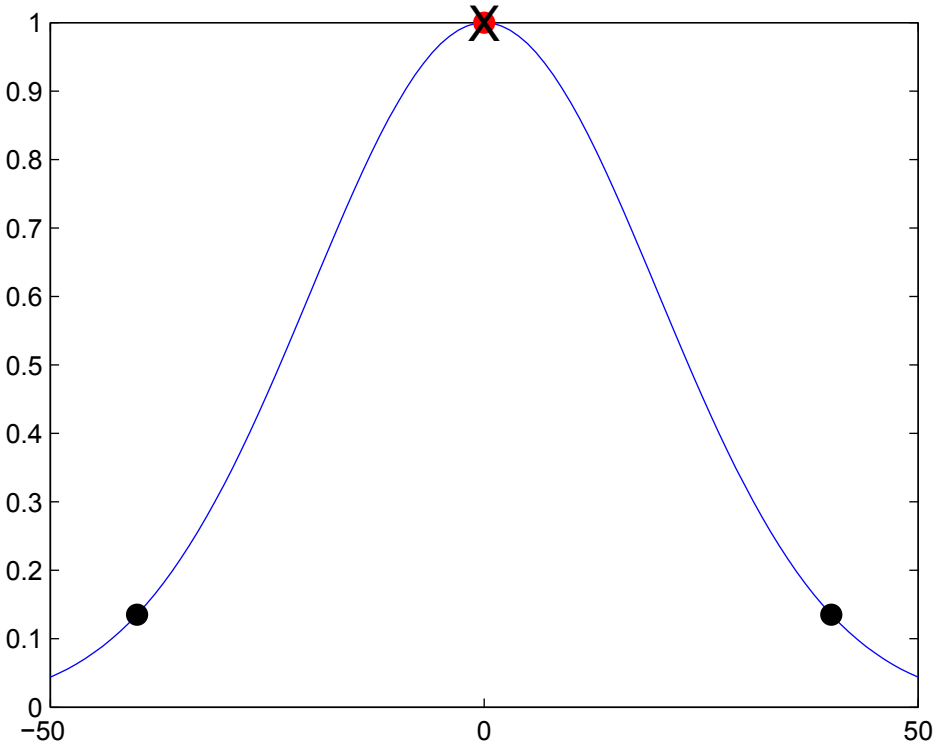
A distância de Hamming computa o número de ocorrências distintas num par de vetores de mesmo comprimento e é uma métrica bastante conhecida por suas aplicações em teoria da informação (DRONGELEN, 2006). Para a aplicação na detecção de ANs, empregamos a abordagem utilizada por Humphries (2011). Os trens correspondentes aos neurônios i e j são divididos em *bins* de duração estabelecida pelo experimentador e convertidos em taxas de disparos t_i e t_j. A seguir, cada *bin* k é classificado segundo a ocorrência ou não de *spikes* nesse intervalo de tempo, recebendo 1 no caso de atividade registrada e 0, caso contrário, independente do número de disparos em cada janela. A métrica computa então a proporção de instantes em que há diferença no comportamento dos trens segundo sua atividade simultânea. A métrica está limitada entre [0,1], tendendo a 0 quanto maior o número de coincidências temporais entre as células no período analisado, conforme a Equação 2.2, a seguir, em que $\oplus$ é o operador "ou exclusivo", N_{bins} corresponde ao número total de *bins* em t_i e t_j e $S_{i,j}$ corresponde ao elemento (i, j) da matriz de similaridade.

$$S_{i,j} = 1 - \frac{\sum_{k=1}^{N_{bins}} (t_i(k) \oplus t_j(k))}{N_{bins}} \tag{2.2}$$

2.2.2 Métricas de similaridade com escala temporal variável

Os métodos descritos a seguir partem de uma ligeira mudança no paradigma da transformação do trem binário para uma escala contínua. Em vez da separação em *bins*, como efetuado para a taxa média de disparo, o vetor pode ser convoluído com diferentes *kernels*

Figura 2.3 – *Kernel* do tipo Gaussiano em torno de um ponto de referência (marcado com um "x"). Os círculos marcados na figura mostram que na vizinhança do ponto de referência a função assume um valor limite (em vermelho) que decai à medida que nos distanciamos (em preto).



Fonte: Produzida pela autora

K de diferentes larguras (σ). Um *kernel* pode ser encarado como uma espécie de medida de similaridade em relação a um ponto de referência, de maneira a assumir valores limites na vizinhança desse ponto de referência e decair a medida que se afasta desse ponto, como representado na Figura 2.3 para um *kernel* do tipo Gaussiano.

Podemos observar que o processamento em *bins* nada mais é do que uma outra versão desse processamento, assumindo um *kernel* do tipo retangular de largura τ numericamente igual à duração temporal de cada *bin*, como pode ser verificado na Equação 2.3.

$$K_\tau(t) = \begin{cases} 1, \ se & \mid t \mid \leq \tau \\ 0, \ se & \mid t \mid > \tau \end{cases} \tag{2.3}$$

É esperado que diferentes *kernels* modifiquem a indicação de dissimilaridade retornada pelas métricas, já que, numa escala temporal adequada eles consistem em maneiras distintas de se traduzir disparos binários para taxas médias de disparo. A largura de um *kernel* (σ) caracteriza sua escala temporal: valores maiores traduzem modificações na taxa de disparo; valores menores, as relações temporais entre os *spikes* (*coincidence detectors*). Uma verificação sistemática do efeito de diferentes *kernels* na detecção de ANs está fora do escopo desse estudo, mas uma análise detalhada comparando o efeito de *kernels* laplacianos, gaussianos, triangulares e retangulares em algumas métricas de similaridade que consideram as relações temporais entre *spikes* gerados segundo o processo de Poisson foi realizada por Paiva *et al.* (2010), que indicou melhorias pouco significativas em modificações de K.

As métricas que utilizam suavizações dessa forma selecionadas para compor esta dissertação foram descritas a seguir.

2.2.2.1 Victor e Purpura

A proposta de Victor e Purpura (1996), uma das primeiras da literatura, estabelece uma métrica baseada no custo necessário para que um trem de disparos possa ser transformado em outro. Para tanto, atribuem-se pesos a três operações básicas: adição, remoção e deslocamento. Um custo fixo unitário é designado às adições e remoções de um disparo e um parâmetro variável (c_v) à movimentação. A métrica computada corresponde à soma das interações. A presença de um parâmetro escolhido pelo experimentador (c_v) atua de maneira a permitir um ajuste da escala temporal de interesse, pois para um baixo valor absoluto de c_v, a métrica converge para a diferença no número de disparos e, para um valor elevado, para a contagem do números de *spikes* não coincidentes. Para valores intermediários de c_v a distância entre dois trens será pequena se eles contiverem um número parecido de *spikes* ocorrendo em instantes de tempo parecidos (VICTOR, 2009).

Como mostrado por Paiva *et al.* (2010), essa métrica pode também ser encarada como um *kernel* triangular invertido, segundo a equação a seguir:

$$K_\sigma(t) = \begin{cases} \frac{1}{\sigma}|t_i^k - t_j^l|, \ se \quad |t_i^k - t_j^l| < 2\sigma \\ 2, \ caso\ contrário \end{cases} \tag{2.4}$$

onde $\sigma = \frac{1}{c_v}$ e t_m^x corresponde ao tempo de ocorrência do *spike* x no neurônio m.

2.2.2.2 Convolução com uma janela e Cosseno

Humphries (2011) utiliza em seu trabalho o uso de janelas (*kernels*) do tipo gaussianas (Equação 2.5 e originalmente proposta por Schreiber *et al.* (2003)) e exponenciais (Equação 2.6) de largura $\sigma = \frac{\delta_{t_i}}{\sqrt{12}}$, onde δ_{t_i} é baseado nos valores do ISI da população.

$$K_\sigma(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-t^2}{2\sigma^2}} \tag{2.5}$$

$$K_\sigma(t) = e^{\frac{-t}{\sigma}} u(t), \text{ onde } u(t) \text{ é a função degrau de Heaviside} \tag{2.6}$$

O uso de *kernels* gaussianos se motiva num janelamento mais suave que o realizado com janelas quadradas, enquanto que o exponencial se inspira no formato de onda aproximado para as correntes pós-sinápticas (DAYAN; ABBOTT, 2001).

Após a convolução, a dissimilaridade é mensurada pelo valor do cosseno dos dois vetores $v_i(t)$ e $v_j(t)$, segundo a Equação 2.7, em que $v_k(t)$ é o valor obtido pela convolução do trem binário k de duração t com o *kernel* e $S_{i,j}$ corresponde ao elemento (i, j) da matriz de similaridade, montada ao computarmos a métrica em relação à células i e j.

$$S_{i,j} = 1 - \frac{v_i(t) v_j(t)}{\|v_i(t)\| \|v_j(t)\|} \tag{2.7}$$

2.2.2.3 Van Rossum

A métrica introduzida por Rossum (2001) também tem como etapa inicial a convolução dos trens com *kernels* exponenciais (Equação 2.6), no entanto, sua largura permanece como um parâmetro de entrada a ser especificado pelo usuário, dependendo da escala temporal de interesse. Valores elevados de σ tornam a métrica sensível à taxa média de disparos, enquanto que valores baixos de σ tornam a métrica sensível à relação temporal entre *spikes*.

A métrica é calculada segundo a Equação 2.8, em que $v_k(t)$ é o valor obtido pela convolução do trem binário k de duração t com o *kernel* e $S_{i,j}$ corresponde ao elemento (i, j) da matriz de similaridade, montada ao computarmos a métrica em relação à célula i e a célula j.

$$S_{i,j} = \frac{1}{\sigma} \int_0^\infty [\mathrm{v_i(t)} - \mathrm{v_j(t)}]^2 \mathrm{d}t \tag{2.8}$$

Essa métrica pode ser interpretada como uma estimativa do efeito de dois trens pré-sinápticos em uma célula pós- sináptica, já que, conforme mencionado na Seção 2.2.2.2,

biologicamente, o formato da corrente pós sináptica é aproximadamente uma exponencial (KREUZ *et al.*, 2007).

2.2.2.4 Billeh

Um dos método mais recentes, proposto por Billeh *et al.* (2014), busca quantificar as interações levando em conta não apenas as características biológicas de um conjunto de células, mas também realiza um processamento diferenciado de acordo com a natureza das interações, se excitação ou inibição, e sua diretividade, apontando qual neurônio tem atividade precedente. Analogamente às métricas já descritas nas Seções 2.2.2.2 e 2.2.2.3, essa proposta também modela o acoplamento entre células segundo uma curva exponencial, visto a semelhança que essa função apresenta com as correntes sinápticas. No entanto, o algoritmo não identifica automaticamente se uma célula apresenta interações de excitação ou inibição, e essa informação deve ser fornecida pelo experimentador. Como padrão, consideram-se interações excitatórias.

A primeira etapa do processamento consiste na criação de uma função que descreve a corrente pós sináptica gerada pelos disparos de um neurônio k, $f_k(t)$. Essa função é composta pela convolução do último disparo do neurônio k (t^k_{last}) com o *kernel* $K_{\sigma_{exc}}(t)$ no caso de uma sinapse excitatória e $1 - K_{\sigma_{ini}}(t)$, para uma inibitória, sendo $K_\sigma(t)$ descrito pela Equação 2.5, resultando na Equação 2.9 a seguir.

$$f_k(t) = \begin{cases} K_{\sigma_{exc}}(t - t^k_{last}) \\ 1 - K_{\sigma_{ini}}(t - t^k_{last}) \end{cases} \tag{2.9}$$

A largura dos *kernels* (σ_{exc} e σ_{ini}) pode ser ajustada de acordo com as características dos dados em análise, porém, por simplicidade e de maneira análoga à proposta do autor, nesse trabalho as larguras serão consideradas iguais ($\sigma_{exc} = \sigma_{ini} = \sigma$).

A seguir, a influência de uma célula i em outra j é computada ao somarmos efeito da corrente pós sináptica de i ($f_i(t)$) nos instantes de tempo em que j dispara, de acordo com a Equação 2.10, em que N_k é o número total de disparos do neurônio k.

$$\sum_{x=1}^{N_j} f_i(t^j_x) \tag{2.10}$$

Esse resultado é então normalizado para a janela de tempo em análise de duração T

e o acoplamento entre as células é calculado para a amostra segundo 2.11b.

$$\begin{cases} \tilde{f}_k(t) = \frac{f_k(t) - < f_k >}{1 - < f_k >} \quad sendo \\ < f_k >= \frac{\int_0^T f_k(t) \mathrm{d}t}{T} \end{cases} \tag{2.11a}$$

$$F_{i,j} = \frac{\sum_{x=1}^{N_j} \tilde{f}_i(t_x^j)}{max(N_i, N_j)} \tag{2.11b}$$

Finalmente, a métrica de similaridade é definida apenas para valores positivos, retornando a matriz S segundo 2.12:

$$S_{i,j} = max(F_{i,j}, 0) \tag{2.12}$$

A definição acima permite que as matrizes de similaridade sejam assimétricas, diferentemente de todas as listadas nessa seção, e possibilita inferências sobre a direcionalidade das interações.

2.2.3 Métricas de similaridade baseadas nas relações temporais entre *spikes*

Finalmente, exploraremos métricas segundo o terceiro e último paradigma de processamento de trens. Como abordado na Seção 1.3, evidências apontaram que o código neuronal baseado apenas em variações da taxa média de disparos ignoravam outras formas de relações entre *spikes*, possivelmente descartando informações sobre os estímulos. Isso foi corroborado por experimentos como o de MacLeod *et al.* (1998), que utilizou as relações temporais para distinguir a atividade referente a dois odores distintos; muitos outros estudos já evidenciaram a importância das relações temporais no processamento de informação no cérebro (PANZERI *et al.*, 2010; GANSEL; SINGER, 2012; BILLEH *et al.*, 2014). Além disso, a probabilidade de que as coincidências temporais entre disparos ocorra nas bordas das janelas empregadas numa transformação em taxa média de disparos é alta, podendo acarretar no descarte de um elevado número de coincidências (GRU; GRüN, 2009). Ademais, pequenas variações no tamanho do *bin* produzem diferenças significativas nos valores das correlações (KRUSKAL *et al.*, 2007). As abordagens selecionadas foram introduzidas a seguir.

2.2.3.1 Sincronização de evento

Introduzida por Quiroga *et al.* (2002), esta métrica quantifica a interação entre duas séries temporais de duração T com base na ocorrência simultânea de eventos; para o caso de análise de trens de disparo, são definidos como a ocorrência de cada potencial de ação.

Os eventos são coincidentes se ocorrerem ao mesmo tempo nos neurônios i e j, mas caso sua diferença esteja dentro de uma pequena janela temporal τ, essa interação também é contabilizada segundo os valores $J_{i,j}$ da Equação 2.13c. Seu valor é acumulado ao longo do tempo na função $c(i|j)$, Equação 2.13b a seguir, onde m_i e m_j representam respectivamente o número total de disparos dos neurônios i e j. Dessa maneira, τ introduz uma robustez ao efeito de perturbações espúrias no potencial de membrana, que podem acarretar em ligeiras diferenças temporais na ocorrência de disparos. A fim de tornar a medida adaptativa à taxa de disparos instantânea e independente da definição de um parâmetro, τ se adapta dinamicamente ao longo do tempo, segundo a Equação 2.13a.

$$\tau_{i,j} = \frac{min(t^i_{k+1} - t^i_k, t^i_k - t^i_{k-1}, t^j_{l+1} - t^j_l, t^j_l - t^j_{l-1})}{2} \tag{2.13a}$$

$$c(i|j) = \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} J_{k,l} \quad onde \tag{2.13b}$$

$$J_{k,l} = \begin{cases} 1, \ se \quad 0 < |t^i_k - t^j_l| \leq \tau_{k,l} \\ 0.5, \ se \quad t^i_k = t^j_l \\ 0, \ caso\ contrário \end{cases} \tag{2.13c}$$

Assim, a matriz de similaridade S é composta por uma função simétrica, segundo a Equação 2.14.

$$S_{i,j} = \frac{c(i|j) + c(j|i)}{\sqrt{m_i m_j}} - 1 \tag{2.14}$$

2.2.3.2 Intervalo entre *spikes* (ISI)

O primeiro passo da métrica proposta por Kreuz *et al.* (2007) consiste num pré-processamento que transforma os trens binários em vetores que correspondem à taxa de disparos instantânea por meio do cálculo do ISI (ν_{ISI}). Para uma célula selecionada n, e para cada instante amostrado t, são localizadas as primeiras ocorrências de um potencial de ação imediatamente antes (t^n_P) e imediatamente depois (t^n_F) de t. Uma ocorrência é inserida artificialmente tanto no começo quanto no final do trem. Assim, para o dado avaliado, de duração total T, temos (KREUZ, 2011):

$$\nu^n_{ISI}(t) = t^n_F(t) - t^n_P(t) \tag{2.15}$$

Então, calcula-se a razão entre a taxa de disparos instantânea dos dois neurônios ν^i_{ISI} e ν^j_{ISI} segundo a equação abaixo:

$$I(t) = \begin{cases} \frac{\nu^i_{ISI}(t)}{\nu^j_{ISI}(t)} - 1, \ se \quad \nu^i_{ISI}(t) \leq \nu^j_{ISI}(t) \\ -\left(\frac{\nu^j_{ISI}(t)}{\nu^x_{ISI}(t)} - 1\right), \ caso\ contrário \end{cases} \tag{2.16}$$

A matriz de similaridade S é então computada pela integração do valor absoluto de $I(t)$:

$$S_{i,j} = \frac{1}{T} \int_{t=0}^{T} |I(t)| \mathrm{d}t \tag{2.17}$$

Sua utilização tem vantagens, por ser não paramétrica, com uma escala temporal intrínseca e adaptativa. No entanto como seu foco é em quantificar similaridades entre taxas de disparos instantâneas, ela tem limitações para distinguir situações em que essa grandeza é constante ao longo do tempo para trens distintos. Por exemplo, a métrica retorna o mesmo valor para trens idênticos e para trens periódicos e de mesmo período (KREUZ *et al.*, 2011).

2.2.3.3 SPIKE

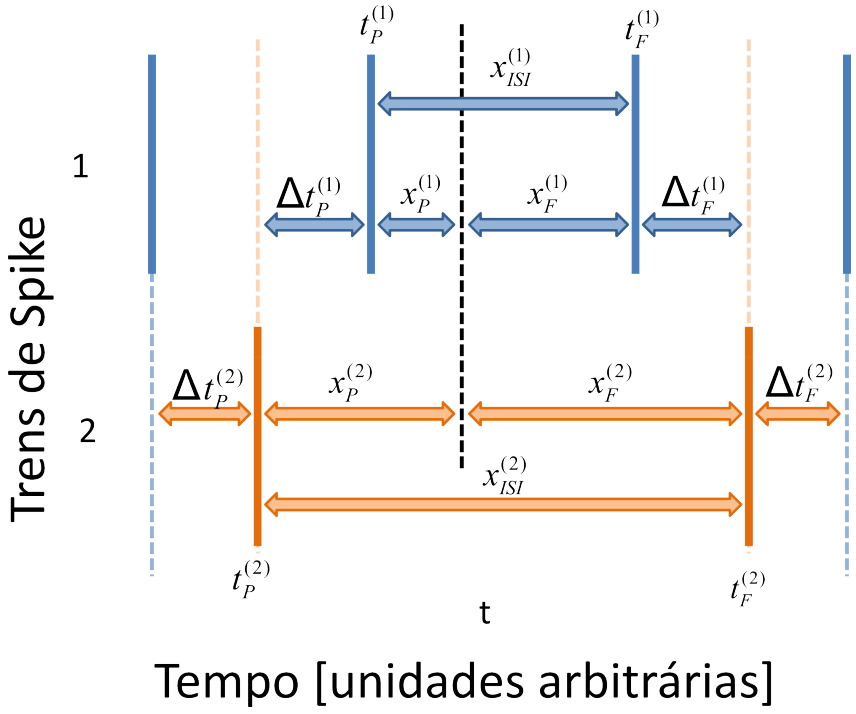
Numa melhoria da métrica apresentada na seção anterior, Kreuz *et al.* (2013) buscou incorporar em seu índice medidas de quantificação de alterações na sincronia temporal das células, ou seja, como as alterações nas proporções de *spikes* coincidentes se modifica ao longo do tempo. Ele propôs a métrica SPIKE, uma extensão da métrica ISI, porém levando em consideração as relações entre células distintas. Para descrevê-la, utilizaremos como apoio a Figura 2.4 a seguir.

Inicialmente, quatro intervalos são definidos, constituindo-se nos disparos de duas células i e j imediatamente antes $(t^k_P(t), k = \{i, j\})$ e depois $(t^k_F(t), k = \{i, j\})$ do instante em análise t. A diferença temporal entre esses instantes $(\Delta t_Y(t), Y = \{P, F\})$ é, então, computada segundo a Equação 2.18 abaixo. Assim como para a métrica ISI (Seção 2.2.3.2), insere-se artificialmente um disparo no começo e no final de cada trem.

$$\begin{cases} \Delta t_P(t) = t^i_P(t) - t^j_P(t) \\ \Delta t_F(t) = t^i_F(t) - t^j_F(t) \end{cases} \tag{2.18}$$

A métrica se baseia, numa ponderação computada em cada instante de tempo t entre as distâncias dos trens adjacentes segundo um peso $(x^k_Y(t), k = \{i, j\}, Y = \{P, F, ISI\})$ calculado segundo relações entre intervalos de tempo, como mostra a Equação 2.19, a seguir:

Figura 2.4 – Na figura, as barras contínuas marcam a ocorrência de um *spike* para as células 1, em azul, e 2, em laranja. A barra hachurada em preto marca o instante de tempo de referência, t. As setas ilustram os intervalos de tempo empregados no cálculo da distância SPIKE.



Fonte: Adaptada de Kreuz (2011)

$$\begin{cases} x_P^k(t) = t - t_P^k(t) \\ x_F^k(t) = t_F^k(t) - t \\ x_{ISI}^k(t) = t_F^k(t) - t_P^k(t) \end{cases} \tag{2.19}$$

Assim, a ponderação das diferenças temporais locais é computada segundo a Equação 2.20:

$$S_k(t) = \frac{\Delta t_P^k(t) x_F^k(t) + \Delta t_F^k(t) x_P^k(t)}{x_{ISI}^k(t)} \quad , k = \{i, j\} \tag{2.20}$$

E, finalmente, uma função que reúne a contribuição dos dois trens, incluindo também a diferença relativa entre eles é ponderada de acordo com a Equação 2.21

$$S(t) = \frac{S_i(t) x_{ISI}^j(t) + S_j(t) x_{ISI}^i(t)}{2(x_{ISI}^i(t) + x_{ISI}^j(t))^2} \tag{2.21}$$

Assim, a matriz de similaridade S entre os neurônios i e j pode ser computada de acordo com a Equação 2.22:

$$S_{i,j} = \frac{1}{T} \int_{t=0}^{T} |S(t)| \mathrm{d}t \tag{2.22}$$

O autor propôs algumas adaptações que não serão exploradas nessa dissertação, em que essa métrica pode ser expandida para englobar o comportamento de uma população num índice único e uma adaptação para aplicações em tempo real (KREUZ *et al.*, 2013). Porém, essa métrica possui limitações para lidar com atrasos, que devem ser previamente tratados e corrigidos.

2.2.4 Adaptações

Nesta seção descrevemos as métricas utilizadas na computação da matriz de similaridade S analisadas nesse trabalho. A fim de homogeneizar a escala das matrizes S, elas foram normalizadas por seu maior valor. Além disso, as métricas podem ser divididas em duas categorias de relações entre os objetos: relações de similaridade e relações de distância.

As relações de similaridade se referem à situação em que há similaridade máxima entre um objeto e ele mesmo, ou seja, o elemento $s_{i,j}$ da matriz de similaridade S é igual a 1 quando $i = j$. Dentre as métricas apresentadas nessa seção, esse tipo de relação está presente na correlação (Seção 2.2.1.1).

Já as relações de distância d são definidas segundo o seguinte conjunto propriedades (PAIVA *et al.*, 2010):

- Simetria: $d(i, j) = d(j, i)$;
- Positividade: $d(i, j) \geq 0$;
- Desigualdade Triangular: $d(i, j) \leq d(i, k) + d(k, j)\ \forall k$
- Identidade: $d(i, j) = 0$, se e somente se $i = j$;

Neste trabalho classificaremos também como métricas de *distância* medidas que, a rigor, seriam denominadas *pseudo-distâncias*, por desobedecerem a propriedade de identidade; *semi-distâncias*, por desobedecerem a desigualdade triangular ou *quasi-distâncias*, por desobedecerem a simetria. Por exemplo, a ISI retorna valor nulo não só para trens de *spike*

idênticos, mas também para dois trens com intervalos entre *spikes* constantes com uma diferença temporal, sendo, portanto, classificada como uma *pseudo-distância*. Já a SPIKE pode desobedecer a desigualdade triangular, e é, a rigor, uma *semi-distância* (MULANSKY *et al.*, 2015). Finalmente, a métrica de Billeh pode retornar valores não simétricos, sendo uma *quasi-distância*.

Matrizes de similaridade S com uma relação de similaridade (S_{sim}) foram convertidas em relações de distância (S_{dist}) segundo a Equação 2.23, a seguir. É importante ressaltar que essa equação já recebe matrizes de similaridade normalizadas.

$$S_{sim} = 1 - S_{dist} \tag{2.23}$$

2.2.5 Comentário geral sobre as limitações das métricas de similaridade

Nesta seção descrevemos de maneira resumida as métricas de similaridade utilizadas nesta dissertação para a análise dos trens de *spikes*. Elas foram classificadas em três categorias, com base na escala temporal utilizada para a caracterização das interações: a divisão em janelas e cálculo da taxa de disparo, as métricas com escala temporal variável e as baseadas exclusivamente nas relações temporais entre os disparos. A determinação da escala temporal adequada é uma tarefa não trivial, que requer uma análise do experimentador, já que impacta significativamente nos resultados. Por isso, propostas não paramétricas e de escala temporal adaptativa, como as propostas ISI (Seção 2.2.3.2) e SPIKE (Seção 2.2.3.3), se tornam vantajosas.

Apesar das métricas apresentadas restringirem sua atuação a uma escala temporal pré-determinada, as evidências apontam que a codificação da informação é feita de maneira complementar e multiplexada, ou seja, que a informação sobre um sinal ou estímulo é processada de maneira distribuída e que diferentes características estão representadas em escalas temporais distintas (PANZERI *et al.*, 2010). Kreuz *et al.* (2007) afirma que muitas métricas apresentam uma alta correlação entre si, o que evidencia a necessidade de novas métricas que possam, de maneira simultânea, caracterizar interações em múltiplas janelas de tempo.

Além disso, com a exceção da métrica apresentada na Seção 2.2.2.4, todas as propostas desconsideram a direcionalidade das interações. Essa diretividade é observada em sistemas biológicos, uma vez que as sinapses ocorrem de neurônios pré para os pós sinápticos e, uma simplificação que desconsidera essa efeito pode descartar informação. Pretendemos com essa dissertação avaliar o impacto dessa consideração nos algoritmos de detecção de comunidades.

Ademais, a maioria dessas métricas foi proposta considerando-se cenários em que a in-

teração entre células ocorre por padrões de excitação mas pouco exploraram sistematicamente casos com inibições ou atrasos. Esses padrões também se relacionam com interações hebbianas (SEJNOWSKI *et al.*, 1989; FRéGNAC, 2003) e estão presentes em circuitos córtico-corticais e córtico-subcorticais (YANG *et al.*, 2012; XU *et al.*, 2012; WANG *et al.*, 2012). Métricas baseadas em excitações geralmente modelam o comportamento das células para a detecção de coincidências (*coincident detectors*), e apontam uma maior similaridade quanto maior o número de coincidências. No entanto, observações experimentais mostram que, na realidade, a imensa maioria das células tem correlações de baixa ordem de magnitude (LONG; CARMENA, 2011).

Finalmente, muitas métricas propostas tem um elevado tempo de processamento e são não causais, ou seja, empregam cálculos em que o estado presente é computado levando-se em conta informações sobre estados futuros. Esses fatores inviabilizam um processamento em tempo real.

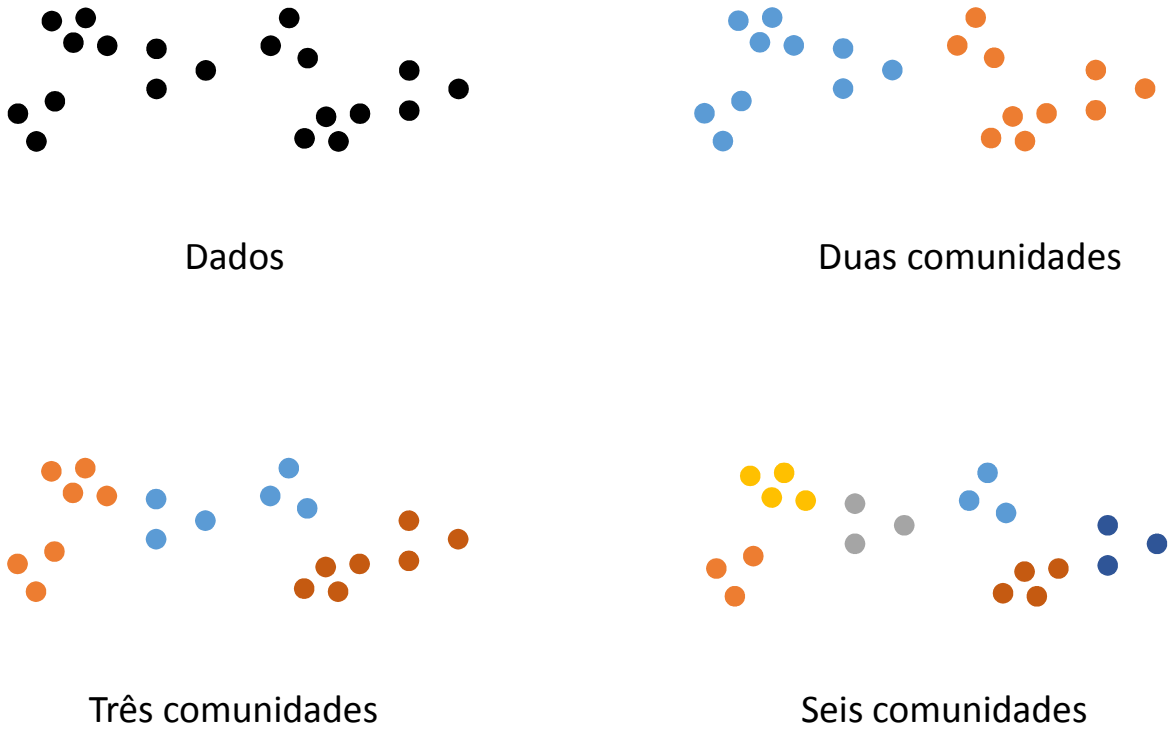
Na próxima seção descrevemos os métodos empregados na detecção de comunidades, que receberam como entrada a matriz de similaridade S.

2.3 Métodos de classificação

Nesta seção, descreveremos os algoritmos de detecção de comunidades que serão avaliados nessa dissertação. Eles receberão como entrada a matriz de similaridade S, computada empregando um dos métodos descritos na seção anterior, e retornarão o número de assembleias encontradas no conjunto de dados, assim como uma identificação de seus membros (i.e., dos neurônios que a compõe). Esses algoritmos tem por objetivo agrupar as células com padrões de disparos mais similares entre si em comparação com as demais. Eles podem estar baseados em noções diferentes do que constitui uma comunidade e qual a maneira mais eficiente de encontrá-las. A escolha do algoritmo mais eficiente permanece um foco de intensa pesquisa em mineração de dados (*data mining*, na literatura em inglês), mas, no momento, é um processo interativo, que envolve vários ajustes de parâmetros e diferentes tentativas até que o resultado obtenha as propriedades desejadas (ESTIVILL-CASTRO, 2002). É importante ressaltar que esse problema não possui uma solução única, já que o número de comunidades detectadas pode depender de vários parâmetros, como por exemplo, a resolução. Esta situação foi representada de maneira esquemática na Figura 2.5.

Este trabalho almeja, portanto, contribuir na exploração de diferentes algoritmos visando a detecção de comunidades em registros eletrofisiológicos extracelulares de trens de *spike*. Neste caso, temos uma correspondência entre as comunidades encontradas e as ANs. Dessa maneira, o foco estará em algoritmos não supervisionados que empregam distintos

Figura 2.5 – Representação esquemática do efeito da resolução na determinação de comunidades.



Fonte: Produzida pela autora

conceitos de agrupamentos e com diferentes níveis de complexidade.

2.3.1 Análise de componentes principais (PCA) e Decomposição em valores singulares (SVD)

A análise de componentes principais (PCA) é uma das técnicas mais conhecidas para processamento multivariável, visando a compressão e a detecção de padrões num conjunto de dados. A ideia é fazer uma transformação dos dados de entrada, selecionando um número reduzido de atributos que ainda seja capaz de caracterizá-los. O PCA consiste na transformação linear ótima segundo o erro quadrático médio e resulta num conjunto de variáveis lineares e descorrelacionadas, as chamadas componentes principais (PCs), que representam as direções que maximizam a taxa de decaimento da variância (HAYKIN, 1999).

O PCA tradicional tem limitações quando lida com dependências de ordem superior, uma vez que é um método linear; é independente da fonte dos dados; não possui parâmetros e, quando aplicado a problemas de classificação, designa cada nova amostra baseado

estritamente no que foi observado durante o período de treinamento, não atualizando dinamicamente suas propriedades estatísticas.

Algumas adaptações foram propostas para lidar com essas limitações, como, por exemplo, o PCA Dinâmico(DPCA)(FALLAHATI *et al.*, 2002) ou o PCA Recursivo (VOEGTLIN, 2005), que lidam com processos não estacionários; o PCA Multi-Escala (BAKSHI, 1998), que visa melhorar o desempenho quando os dados tem escalas temporais diferentes; e, finalmente, o PCA Não-Linear (FOTHERINGHAME; BADDELEY, 1997), e o *Kernel*-PCA (SCHöLKOPF *et al.*, 1998) que visam interações não lineares.

Especificamente para o problema de detecção de ANs, as PCs $\mathbf{p_i}$ são calculadas a partir dos autovetores $P = [\mathbf{p_1}, ..., \mathbf{p_{n_{Neurons}}}]$ da matriz de covariância computada a partir da taxa de disparos (SANTOS *et al.*, 2011; SANTOS *et al.*, 2013), conforme descrito na Seção 2.2.1.1. Detsa maneira, o uso do PCA neste problema específico pressupõe que comunidades serão definidas como conjuntos de células que apresentarem disparos correlacionados entre si.

Nesse trabalho exploraremos também os resultados para matrizes de similaridade com métricas distintas, descritas na Seção 2.2. No entanto, como por definição o PCA emprega ou a matriz de covariância ou a de correlação, denominaremos os casos que utilizam outras métricas de similaridade de decomposição em valores singulares (SVD). Agora as PCs $\mathbf{p_i}$ serão dadas pelos autovetores $P = [\mathbf{p_1}, ..., \mathbf{p_{n_{Neurons}}}]$ da matriz de similaridade S. É importante observar que o paralelo estabelecido entre o SVD e o PCA só é válido quando os autovetores de S são independentes entre si, o que requer que S seja simétrica e, portanto, impõe uma restrição adicional às métricas de similaridade.

Para determinar se um grupo de neurônios apresenta disparos correlacionados, portanto formando uma AN, ou se seus disparos são independentes entre si, empregaremos a proposta de Peyrache *et al.* (2010) e Santos *et al.* (2011), baseada na distribuição de Marčenko-Pastur, porém um limiar determinado por um percentil de dados substitutos, ou *surrogate data*, na literatura em inglês, também pode ser utilizado. Marcenko e Pastur (1967) mostraram que a matriz de autocorrelação calculada para N_{linhas} vetores independentes entre si e de comprimento N_{cols} possui autovalores (λ) segundo a distribuição analítica $p(\lambda)$, descrita pela Equação 2.24, em que $q = \frac{N_{cols}}{N_{linhas}}$ e σ^2 é a variância dos elementos dessa matriz.

$$\begin{cases} p(\lambda) = \frac{q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_{max}-\lambda)(\lambda-\lambda_{min})}}{\lambda} \;\; em\ que \\ \lambda_{min} < \lambda < \lambda_{max};\ e \\ \lambda_{min}^{max} = \sigma^2(1 \pm \sqrt{\frac{1}{q}})^2 \end{cases} \tag{2.24}$$

Portanto, a Equação 2.24 pode ser usada na definição de um limiar mínimo de interação necessária entre um grupo para que ele seja considerado uma assembleia, já que a probabilidade da matriz de autocorrelação de vetores independentes apresentarem um autovalor maior que λ_{max} é nula (SANTOS *et al.*, 2013). Se, por exemplo, um determinado grupo de neurônios formar uma AN, sua atividade estará associada a uma PC e o autovalor correspondente será maior que o limiar λ_{max} definido pela distribuição de Marčenko-Pastur; analogamente, se os neurônios forem independentes, o autovalor λ associado a essa PC estará localizado no interior dessa distribuição ($\lambda_{min} \leq \lambda \leq \lambda_{max}$).

Os pesos de cada PC, dados pelos elementos w_k, $k = 1, ...n_{Neurons}$, de uma PC $\mathbf{p_i}$, podem indicar os neurônios que pertencem a cada AN, uma vez que um padrão característico de disparo para uma assembleia consiste em pesos iguais para seus membros e pesos nulos para os neurônios independentes. Uma situação próxima da ideal é exemplificado nas Figuras 2.6a e 2.6c, compostas por duas assembleias disjuntas. O autovetor correspondente a cada uma delas é exibido em cada uma das figuras, e a observação de $\mathbf{p_1}$ e $\mathbf{p_2}$ permite inferir os membros dessas ANs.

No entanto, existem limitações quando duas assembleias compartilham neurônios: nesse caso, a matriz de covariância ou a matriz de similaridade agrupa o efeito dos padrões de disparos distintos e isso modifica a distribuição dos pesos das PCs (Figuras 2.6b e 2.6d), impossibilitando a identificação dos membros de cada AN (SANTOS *et al.*, 2011; SANTOS *et al.*, 2013).

Para determinar quais neurônios pertencem a uma assembleia, estabelecemos um limiar segundo a Equação 2.25, abaixo. Um neurônio k foi considerado como parte da assembleia apenas no caso em que $|w_k|$ fosse maior que Thr.

$$Thr = G \sum_{k=1}^{n_{Neurons}} \frac{|w_k|}{n_{Neurons}} \tag{2.25}$$

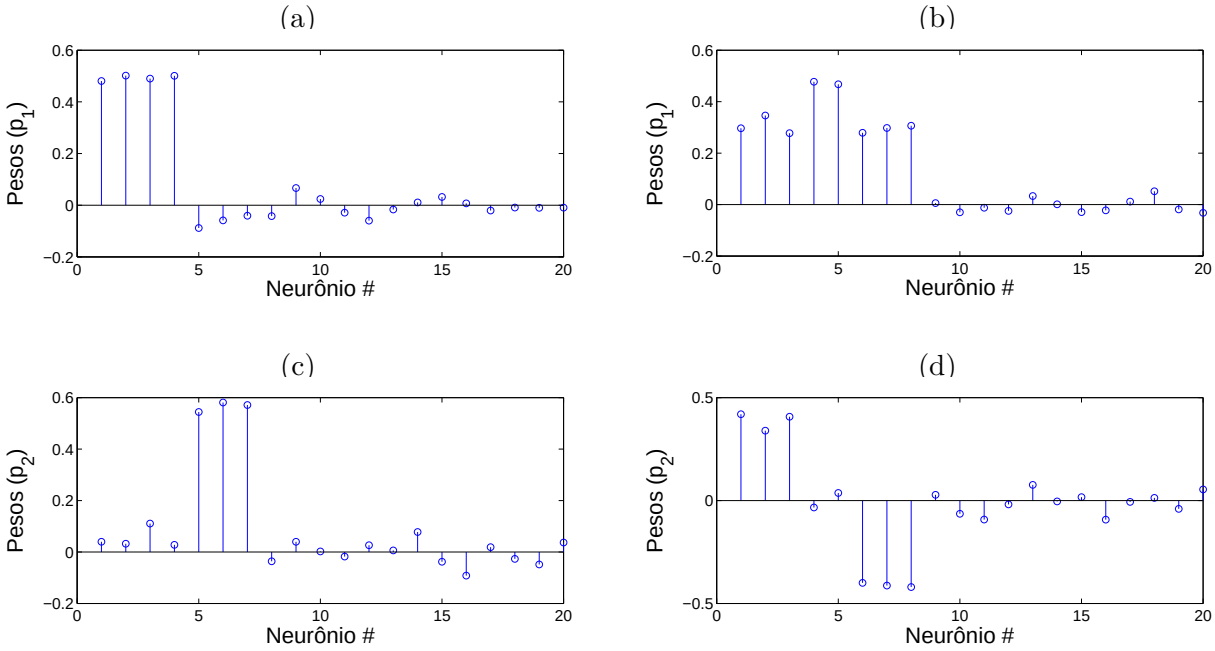
em que $n_{Neurons}$ representa o número total de neurônios da população analisada.

O fator G da Equação 2.25 foi ajustado com base em várias simulações de assembleias disjuntas segundo um modelo poissonico de disparo (Seção 2.1.2), até que a relação entre falsos-positivos e falsos negativos estivesse balanceada.

2.3.2 Análise de componentes independentes (ICA)

O ICA é uma técnica de processamento multivariável que visa a separação de um sinal em fontes estatisticamente independentes entre si, assumindo que o sinal resultante é uma combinação linear dos sinais de cada uma das fontes. O algoritmo se baseia no Teorema

Figura 2.6 – Padrão característico de uma assembleia: os membros apresentam pesos parecidos enquanto que os outros neurônios aparecem com pesos nulos. As Figuras 2.6a e 2.6c apresentam PCs significativas para as ANs disjuntas e exibem um exemplo em que a distribuição dos pesos permite a identificação dos membros. No entanto, as PCs das ANs que compartilham neurônios (Figuras 2.6b e 2.6d) tem uma distribuição não característica e a classificação dos neurônios não pode mais ser determinada.



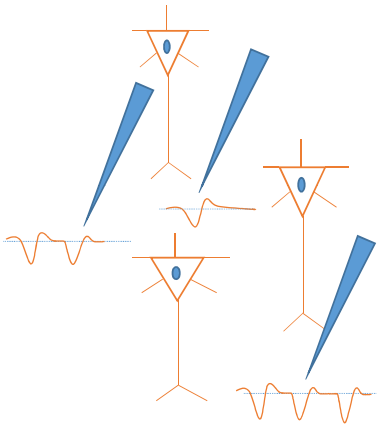
Fonte: Adaptada de Santos *et al.* (2013)

do Limite Central, que enuncia que a soma de variáveis independentes com variâncias finitas e identicamente distribuídas, quaisquer que sejam suas distribuições, converge para uma distribuição gaussiana (PAPOULIS, 1991). O teorema implica que, no caso de fontes não gaussianas, a combinação do sinal de fontes distintas é mais gaussiana que os sinais originais. Esta propriedade é explorada pelo algoritmo, pois ele busca minimizar uma função que mede a gaussianidade do sinal (HYVäRINEN; OJA, 2000; SANTOS *et al.*, 2013).

Existem várias propostas de computação do ICA, e a grande maioria se diferencia basicamente pela definição dessa função que computa a gaussianidade do sinal (HYVäRINEN; OJA, 2000). Utilizaremos nessa dissertação a proposta de Santos *et al.* (2013), que emprega o fastICA, originalmente proposto por Bingham e Hyvärinen (2000).

A análise de componentes independentes (ICA) é empregada na detecção de ANs como um passo complementar ao processamento descrito na seção anterior, visando uma

Figura 2.7 – Representação de uma cenário de aplicação do ICA. Três neurônios são gravados por três eletrodos distintos. O sinal obtido por cada eletrodo é uma combinação do sinal de cada uma das células.



Fonte: Produzida pela autora

solução para o problema ilustrado no conjunto de Figuras 2.6b e 2.6d. Como será explicado a seguir, esse algoritmo atuará na distribuição dos pesos w_k, de maneira que ela se assemelhe ao padrão característico, ou seja, os neurônios que pertencem à AN apresentarão pesos aproximadamente iguais e os independentes, pesos próximos a zero. Iniciaremos a seção com uma descrição geral do algoritmo e, a seguir, detalharemos seu processamento em conjunto com o PCA ou com o SVD.

Como já mencionado na Seção 2.1, a atividade de um neurônio pode ser simplificada por um processo de Poisson, e, portanto, não é gaussiana, o que permite a aplicação do ICA com sucesso para a identificação de fontes independentes. Para ilustrar sua atuação, faremos uma adaptação para o contexto eletrofisiológico de um exemplo muito conhecido de aplicação do ICA, em inglês denominado *cocktail party*.

Imaginemos uma situação hipotética em que três células distintas são gravadas ao longo do tempo por três eletrodos diferentes, como disposto na Figura 2.7. Denotaremos o sinal de cada célula j por $s_j(t)$ e por $x_i(t)$o sinal gravado pelo eletrodo i. Cada eletrodo gravará o potencial elétrico em seu entorno, incapaz de distinguir a presença dos neurônios, portanto, se desconsiderarmos fatores adicionais, podemos modelar o sinal dos eletrodos como uma combinação da atividade dos neurônios, segundo a Equação 2.26 abaixo:

$$\begin{cases} x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \\ x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \\ x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) \end{cases} \tag{2.26}$$

Nesse caso, os pesos a_{ij} serão dependentes da distância do eletrodo i a cada célula j. Para o caso de uma análise de registros eletrofisiológicos extracelulares(trens de *spike*), desconhecemos tanto a matriz $A = [a_{i,j}]$ quanto a natureza das fontes. A aplicação do ICA nesse caso, com fontes não gaussianas e independentes, permite a separação dos sinais $s_j(t)$.

A determinação das fontes independentes motivou o uso de ICA para registros extracelulares de trens de *spike* por Laubach *et al.* (1999) no problema de detecção de ANs. Como nesse contexto o interesse é em fontes associadas a conjuntos de células e não a células independentes, foi proposto um pré-processamento dos dados empregando o PCA ou o SVD, para a redução de dimensionalidade dos dados, como será explicado a seguir.

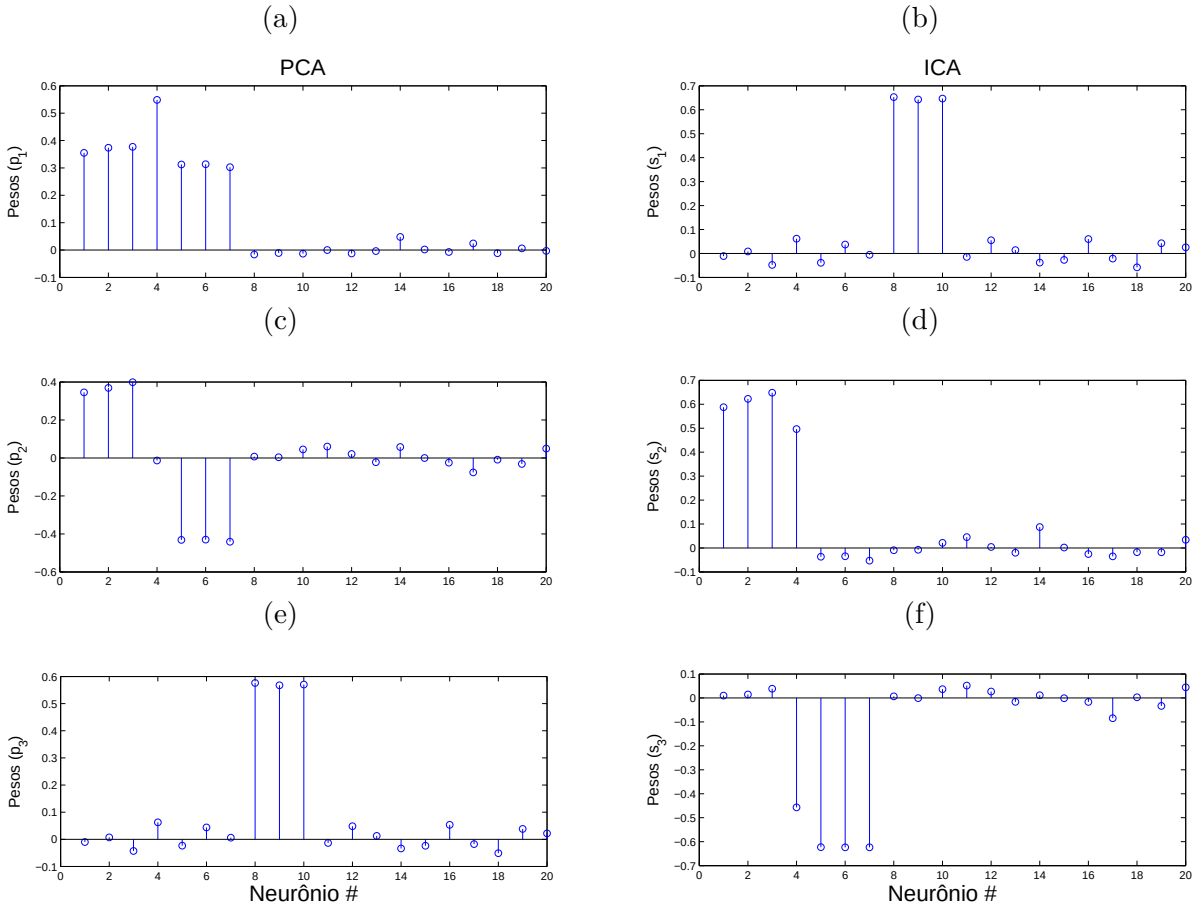
Inicialmente, um número de PCs é selecionado, que corresponderá ao número total de assembleias detectadas *nda* e, a seguir, os dados são projetados no subespaço $T = [\mathbf{p_1}, ..., \mathbf{p_{nda}}]$, definido por esses vetores. Laubach *et al.* (1999), em seu trabalho, rejeitaram todas as PCs associadas a autovalores inferiores a 1. Esse limiar foi aprimorado por Santos *et al.* (2013), que propôs o limiar λ_{max} estabelecido pela distribuição de Marčenko-Pastur (Equação 2.24).

O ICA então recebe como entrada o resultado da projeção dos dados em T e retorna fontes independentes s_i, $i = \{1, ..., nda\}$. É possível observar que os pesos a_k, $k = \{1, ...n_{Neurons}\}$ de s_i estarão distribuídos de maneira próxima ao padrão característico, ou seja, o valor de a_k para os elementos que pertencem à AN será aproximadamente igual e próximo de zero para os que não pertencem.

Assim, o ICA torna possível a distinção das ANs nos casos com neurônios compartilhados, já que agora elas são encaradas como fontes independentes. Na Figura 2.8, é possível verificar para o mesmo conjunto de dados, como a limitação do PCA ou SVD (Figuras 2.8a, 2.8c e 2.8e) aplicado a populações com neurônios em comum é solucionada após a aplicação do ICA (Figuras 2.8b, 2.8d e 2.8f).

A definição da filiação dos neurônios de cada assembleia permanece segundo a proposta da seção anterior, tendo por base a Equação 2.25, porém, substituindo w_k por a_k.

Figura 2.8 – A figura mostra o resultado para a simulação de três ANs, uma independente e as outras duas compartilhando células. Após a aplicação do PCA ou SVD foi possível observar que o número de autovalores maiores que o limite superior da distribuição de Marčenko-Pastur foi igual a três e que, portanto, o número de assembleias foi corretamente identificado. No entanto, o padrão dos pesos das PCs associadas não permite apontar de maneira imediata a filiação das células para as comunidades que compartilham neurônios, Figuras 2.8a, 2.8c e 2.8e. Para o mesmo conjunto de dados, esse problema é solucionado após a aplicação do ICA e os pesos se distribuem aproximadamente segundo o padrão característico, Figuras 2.8b, 2.8d e 2.8f.

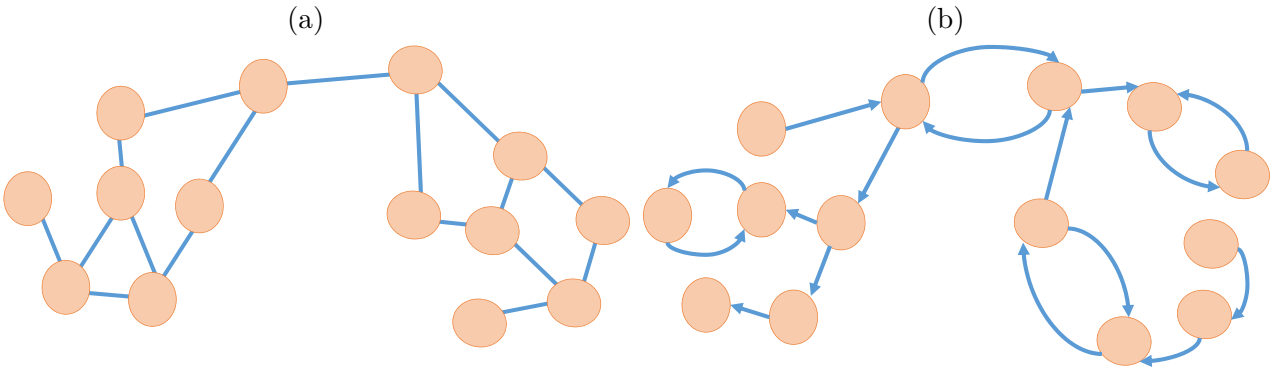


Fonte: Adaptada de Santos *et al.* (2013)

2.3.3 Modularidade

A consideração de ANs como a unidade funcional do sistema nervoso possibilita que redes neurais sejam encaradas segundo um modelo de grafos. Um grafo pode ser encarado como uma representação simplificada de objetos e das relações que eles apresentam entre si. Cada objeto corresponde a um nó e as relações entre eles, denominadas links, são representa-

Figura 2.9 – Representações de uma rede utilizando grafos. Os nós são representados pelos círculos e as interações entre eles, os links, pelos traços, no caso de interações simultâneas e não dirigidas (Figura 2.9a) e por setas no caso de interações direcionadas (Figura 2.9b).



Fonte: Produzida pela autora

das por setas, no caso de interações direcionadas, ou por traços, para interações simultâneas e não dirigidas, como verificado nas Figuras 2.9a e 2.9b.

A Modularidade, o algoritmo descrito nessa seção, foi originalmente proposto por Newman (2006) para o problema de detecção de comunidades segundo o contexto de um ramo da matemática e da computação denominado Teoria de Grafos. Nesse modelo, comunidades são definidas como conjuntos mais densamente conectados em seu interior e com conexões mais esparsas entre si.

O algoritmo visa maximizar uma função de benefício Q (Equação 2.27), denominada Modularidade, que quantifica a diferença entre o número de links dentro das comunidades (α) e o número esperado desses links em redes aleatórias de topologia semelhante (β), com tamanhos equivalentes de nós e links. O máximo da função Q está associado a disposição que maximiza a interação intra-grupos e minimiza a inter-grupos.

$$Q = \alpha - \beta \tag{2.27}$$

Assim, caso Q assuma um valor positivo, a definição de comunidades é satisfeita e, o grafo é particionado em dois subgrupos distintos, que podem ser compostos por uma ou mais comunidades. O algoritmo é executado múltiplas vezes para cada subgrupo até o caso limite, em que $Q \leq 0$.

Como a partição é feita sempre aos pares, podemos analisar o caso de uma interação, sem perda de generalidade. A rede pode ser descrita segundo um vetor z tal que, cada nó z_i

possa assumir valores 1 e -1, referenciando a comunidade a que esse nó pertencerá. Determinar z é o objetivo do algoritmo. As interações da rede podem ser representadas por uma matriz A, denominada matriz de adjacência, de maneira que cada elemento $a_{i,j}$ receba o valor 1 no caso do nó i estar conectado ao nó j e 0 caso contrário. A proposta original de Newman (2006) faz uma simplificação e considera apenas redes não direcionadas, o que implica uma matriz A simétrica. Dessa maneira, α pode ser computado segundo a Equação 2.28

$$\alpha = z^T A z \tag{2.28}$$

β, por sua vez, é calculado segundo um modelo matemático probabilístico, denominado modelo nulo. O modelo gera uma matriz P, na qual computa-se a probabilidade que um nó escolhido aleatoriamente i possua um link com outro nó j. Várias propostas para esse modelo podem ser utilizadas, sendo a empregada originalmente definida pela Equação 2.29, a seguir, em que k_v corresponde ao número de links conectados ao nó v e m ao número total de links na população.

$$P_{i,j} = \frac{k_i k_j}{2m} \tag{2.29}$$

Assim, de maneira análoga, β é calculado de acordo com a Equação 2.30

$$\beta = z^T P z \tag{2.30}$$

Substituindo as equações 2.28 e 2.30 em 2.27, podemos definir uma matriz de modularidade, B:

$$\begin{cases} Q = z^T(A - P)z = z^T B z \;\; sendo \\ B = A - P \end{cases} \tag{2.31}$$

Newman (2006) mostrou que os autovalores positivos de B correspondem a nós mais conectados entre si que o esperado e que o inverso ocorre para valores negativos. Além disso, ele expôs que o maior valor de comunidades presentes em uma rede M pode ser determinado pela quantidade de autovalores positivos η, adicionados de 1. Ele propôs então que o autovetor $u_{\lambda_{MAX}}$ correspondente ao maior autovalor λ_{MAX} de B poderia ser utilizado na determinação de z, em que para cada nó i, atribui-se um índice segundo a Equação 2.32 :

$$z_i = \begin{cases} 1 \;\; se \; u_i \geq 0 \\ -1 \;\; se \; u_i < 0 \end{cases} \tag{2.32}$$

Humphries (2011) adaptou e aprimorou o algoritmo de Newman (2006) para o contexto de detecção de ANs. Ele propôs uma expansão que faz a divisão em comunidades que usa não apenas o maior, mas todos os autovalores positivos de B e emprega, no lugar da matriz de adjacência A, uma matriz computada segundo uma métrica de similaridade S conforme as descritas na Seção 2.2.

Assim, algumas modificações foram estabelecidas: Num caso em que a população fosse constituída por um número N de comunidades, o conceito do vetor z é expandido para uma matriz S, segundo a Equação 2.33 abaixo:

$$Z_{i,n} = \begin{cases} 1 \;\; se\ o\ nó\ i\ está\ na\ comunidade\ n \\ 0\ caso\ contrário \end{cases} \tag{2.33}$$

E a Modularidade fica reescrita segundo a equação:

$$\begin{cases} Q = Tr(Z^T BZ) \;\; em\ que\ B\ é\ redefinida\ por: \\ B = S - P \end{cases} \tag{2.34}$$

A partir dessa nova definição de B, Humphries (2011) separou todos os η autovalores positivos, determinando o maior valor possível de comunidades M. A seguir, esta informação foi utilizada para inicializar um algoritmo de agrupamento denominado *K-means*, que visa particionar um conjunto de dados em um número de agrupamentos determinado por um parâmetro de entrada. O algoritmo *K-means* é dependente de seus parâmetros de inicialização, e, para trazer robustez ao resultado, Humphries (2011) o executou 20 vezes para cada valor possível de comunidades $N = \{2, ..., M\}$. O valor de Q foi armazenado para cada uma dessas interações, e a matriz Z correspondente ao máximo Q foi utilizada para definir as ANs presentes no conjunto de dados de entrada.

2.3.4 Análise do agrupamento hierárquico

Uma condição necessária para a utilização de algoritmos de agrupamento hierárquico é a de que as métricas de similaridade, que expressam as relações entre dois objetos i e j, possam ser interpretadas como uma métrica de distância d, e, portanto, que obedeçam as propriedades descritas na Seção 2.2.4.

Dessa maneira, uma comunidade é formada por objetos mais próximos entre si segundo d, pois eles estarão mais relacionados do que objetos distantes. Vale lembrar que, devido às propriedades de uma métrica de distância, S será uma matriz simétrica. Portanto, em alguns cenários, o uso da métrica de Billeh *et al.* (2014) (Seção 2.2.2.4) não é apropriado. Houghton e Victor (2010) afirmou que algoritmos de agrupamento baseados em métricas de

distância também apresentam bons resultados quando associados a semi-distâncias, portanto, estas métricas serão incluídas. Para as pseudo-distâncias, trabalharemos com um espaço equivalente, que converte o espaço de uma pseudo-distância num espaço completo, para que a relação seja então uma métrica de distância. Por exemplo, para a métrica ISI, seguiremos a sugestão de Mulansky *et al.* (2015) e agruparemos em uma classe equivalente todos os conjuntos de trens de *spike* com o mesmo intervalo entre disparos, mas deslocados no tempo.

A análise do agrupamento hierárquico se inicia pela construção do chamado dendrograma (Figuras 2.10, 2.12, 2.15), que nada mais é do que uma representação visual da matriz de similaridade S. O primeiro passo consiste em encontrar o elemento $S_{x,y} = S_{y,x}$, correspondendo aos objetos X e Y, com a menor distância segundo d.

Esses objetos serão agrupados no par (X, Y) e uma nova matriz precisará ser computada. Precisamos, então, definir uma maneira de quantificar a relação desse novo par com os outros objetos. Essa quantificação (D), denominada *Linkage* na literatura em inglês, pode ser definida de várias maneiras diferentes. Kreuz *et al.* (2007) utilizou em seu trabalho para detecção de ANs a ligação simples. Neste trabalho testaremos também outras formas de quantificação: a completa, a média simples e a média ponderada. Detalharemos cada uma delas nas próximas Seções.

Para ilustrar cada uma das formas de ligação, utilizaremos a seguinte matriz de similaridade calculada a partir de um dado conjunto de objetos:

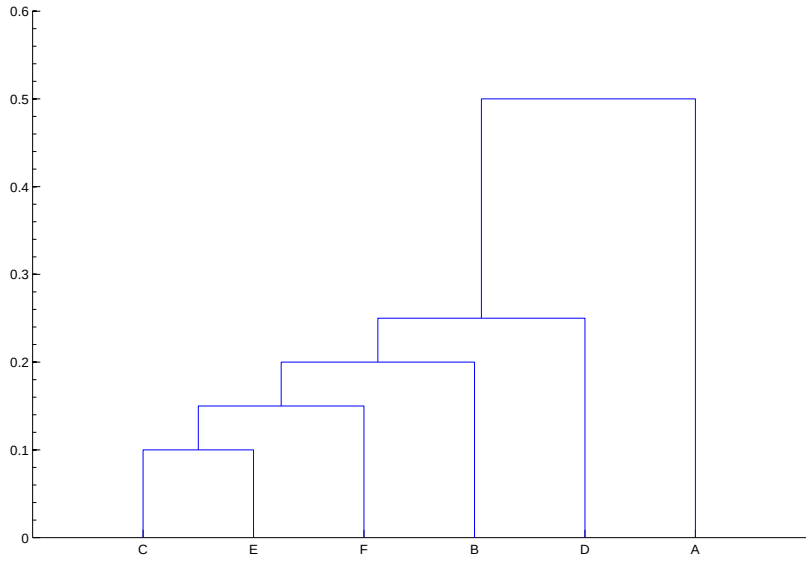
objetos	A	B	C	D	E	F
A	0	0.50	0.65	0.8	1	0.55
B	0.50	0	0.20	0.35	0.9	0.7
C	0.65	0.20	0	0.25	0.1	0.4
D	0.8	0.35	0.25	0	0.7	0.65
E	1	0.9	0.1	0.7	0	0.15
F	0.55	0.7	0.4	0.65	0.15	0

2.3.4.1 Ligação Simples

A comparação do par (X, Y) com um objeto Z segundo a ligação simples é definida como a menor distância entre eles, ou seja, $D((X, Y), Z) = min(d(X, Z), d(Y, Z))$. Esse processo é repetido até o completo agrupamento da matriz.

Para a matriz de similaridade utilizada como exemplo, temos que o menor valor encontrado é para o par (C,E). Dessa maneira, esse par será agrupado por um traço de

Figura 2.10 – Exemplo de dendrograma, gerado a partir da matriz de similaridade descrita na Seção 2.3.4 e utilizando o método de ligação simples.



Fonte: Produzida pela autora

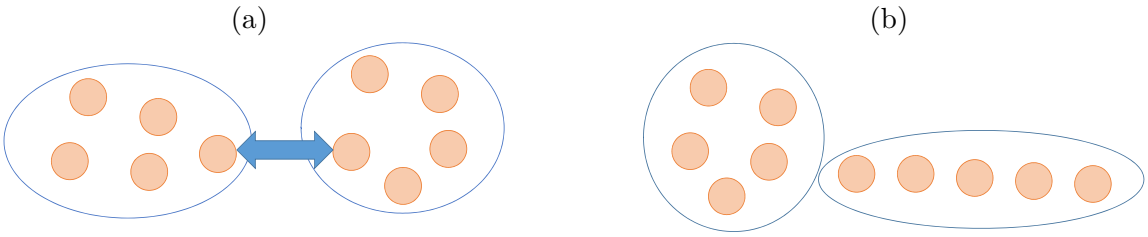
altura 0.1 (Figura 2.10) e uma nova matriz será calculada. Observamos que a distância entre A e C é 0.65 e a entre A e E é 1. Aplicando a equação da ligação simples, temos que a distância entre A e (C,E) será dada por 0.65. Esse processamento se repetirá para os demais objetos, que resultará em uma nova matriz:

$objetos$	A	B	(C,E)	D	F
A	0	0.50	0.65	0.8	0.55
B	0.50	0	0.20	0.35	0.7
(C,E)	0.65	0.20	0	0.25	0.15
D	0.8	0.35	0.25	0	0.65
F	0.55	0.7	0.15	0.65	0

Podemos notar que o próximo agrupamento será entre (C,E,F). Continuando o procedimento, chegaremos ao dendrograma que pode ser visualizado na Figura 2.10:

O uso da ligação do tipo simples é adequado quando a separação entre as comunidades é clara (Figura 2.11a). No entanto, esse tipo de metodologia pode agrupar objetos de uma

Figura 2.11 – Representação esquemática da ligação do tipo simples. A distância entre comunidades é calculada como a menor distância entre membros (Figura 2.11a). Agrupamento distorcido, que pode conter objetos heterogêneos entre si (Figura 2.11b).



Fonte: Produzida pela autora.

forma distorcida e em cadeia, adicionando um objeto após o outro ao agrupamento, ainda que a distância entre o primeiro e o último elementos seja elevada, como esquematizado na Figura 2.11b. Isso pode levar à demarcação de uma comunidade muito heterogênea entre si.

2.3.4.2 Ligação Completa

Para a ligação completa, por sua vez, a comparação do par (X, Y) com um objeto Z é definida como a maior distância entre eles, segundo a expressão $D((X, Y), Z) = max(d(X, Z), d(Y, Z))$. Utilizando a mesma matriz de similaridade apresentada na Seção 2.3.4, teremos um novo dendrograma, exibido na Figura 2.12.

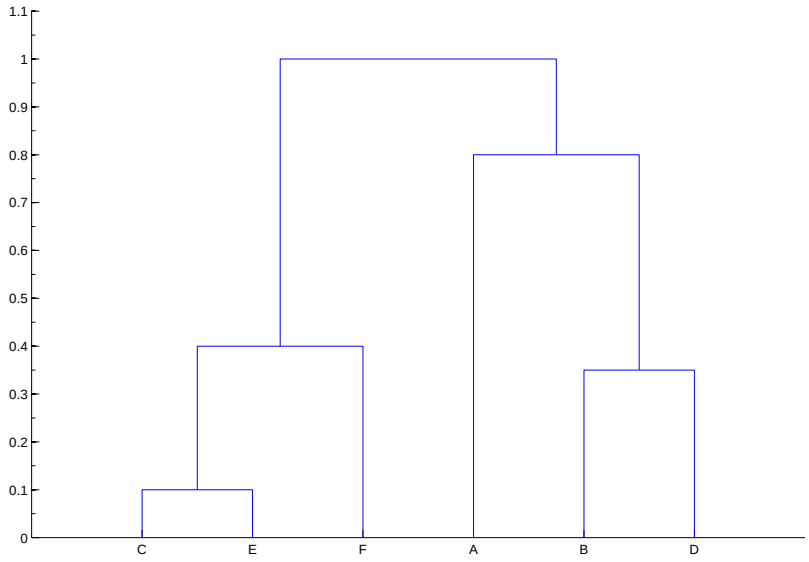
Uma vez que se baseia na maior distância entre um objeto de uma comunidade e um objeto de outra comunidade (Figura 2.13), a ligação do tipo completa é sensível a *outliers*. Além disso, o método de agrupamento tende a separar os objetos em comunidades de tamanhos similares entre si.

2.3.4.3 Ligação Média

Em contrapartida às ligações simples e completa, na ligação média a comparação entre duas comunidades passa a passa a ser calculada a partir de medidas mais centrais. Isso pode ser efetuado de duas maneiras, ao utilizarmos a ligação média simples e a ligação média ponderada. Uma representação é apresentada na Figura 2.14.

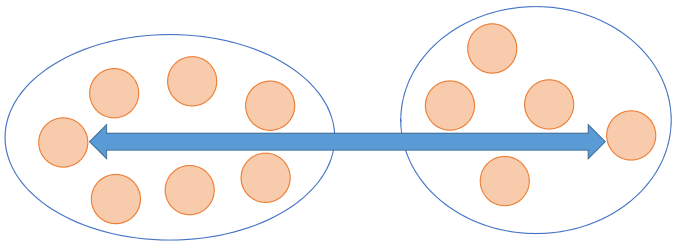
Na ligação média simples, quantifica-se a interação entre duas comunidades calculando-se a distância média entre todos os elementos que as compõe, ou seja, se uma comunidade R, composta por n_r objetos $x_i, i \in (1, ..., n_r)$ e outra comunidade S, com n_s objetos $y_j, j \in (1, ..., n_j)$, então $D(R, S) = \frac{1}{n_i n_j} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} d(x_i, y_j)$.

Figura 2.12 – Exemplo de dendrograma, gerado a partir da matriz de similaridade descrita na Seção 2.3.4 e utilizando o método de ligação completa.



Fonte: Produzida pela autora

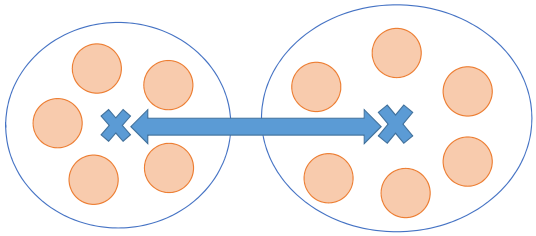
Figura 2.13 – Representação esquemática da ligação completa, em que a distância entre duas comunidades é computada em relação aos objetos mais distantes entre si. Dessa forma, o agrupamento é sensível a *outliers*.



Fonte: Produzida pela autora

Alternativamente, na ligação média ponderada a interação entre comunidades é redefinida como a média das distâncias a medida que novos objetos são incorporados, ou seja, $D((X,Y),Z) = \frac{d(X,Z)+d(Y,Z)}{2}$. Os dendrogramas gerados para a matriz de exemplo para as ligações médias simples e poderada podem ser visualizado nas Figuras 2.15a e 2.15b, respectivamente.

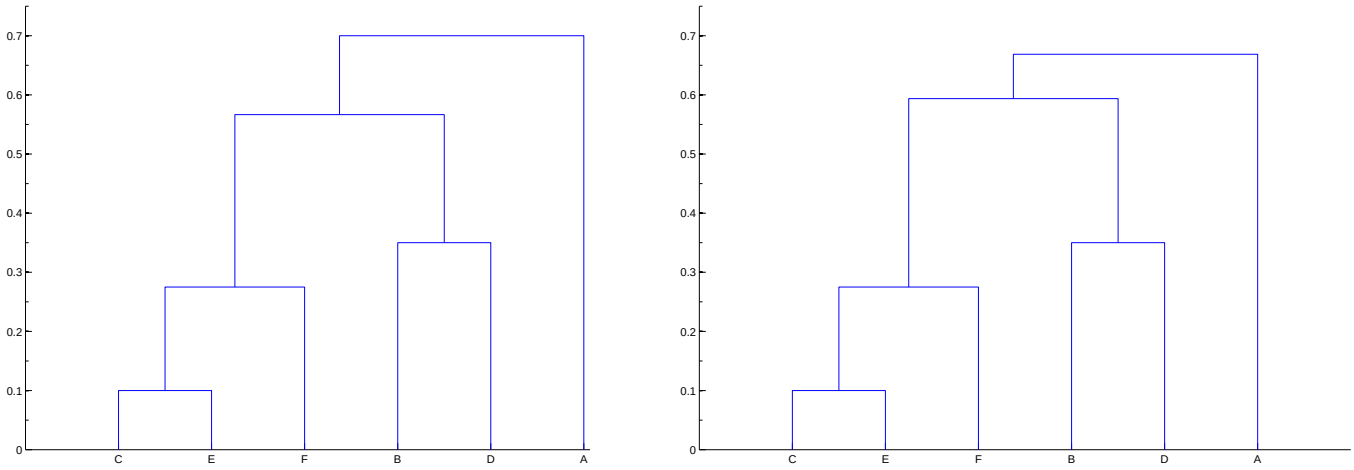
Figura 2.14 – Representação esquemática da ligação média, em que a distância entre duas comunidades é computada a partir da média das distâncias entre os objetos que as compõe (Marcada por um 'X').



Fonte: Produzida pela autora

Figura 2.15 – Dendrogramas gerados a partir da matriz de similaridade descrita na Seção 2.3.4, utilizando as ligações médias simples (Figura 2.15a)e ponderada (Figura 2.15b).

(a) (b)



Fonte: Produzida pela autora.

Esses tipos de ligação se tornam interessantes já que homogeneizam a interação de comunidades por seu tamanho e, portanto, lidam um pouco melhor com comunidades compostas por poucos objetos.

O próximo passo consiste em caracterizar de maneira automática, a partir da análise do dendrograma, a presença de comunidades. Conforme já mencionado na Seção 2.3, esse

problema não é de solução única. Uma das soluções é a definição de um limiar de distância, ou seja, um valor para o eixo vertical das Figuras 2.10, 2.12 e 2.15 a partir do qual todos as ligações com valores inferiores sejam consideras comunidades. Outra opção é a definição de uma distância relativa entre os links que, se excedida, acusará agrupamentos distintos. No contexto de detecção de ANs, Kreuz *et al.* (2007) criaram um índice que computa um peso relativo, porém, essa expressão foi confeccionada para o caso específico de três comunidades e não pode ser generalizada.

No presente trabalho pretendíamos explorar maneiras automática de particionar dendrogramas e identificar ANs, tomando por base simulações biologicamente inspiradas. Assim, a abordagem escolhida consistiu no particionamento sistemático do dendrograma num número de comunidades variando de duas até o número total de objetos. Cada partição foi avaliada segundo índices de validação, descritos na Seção 2.3.4.4. O particionamento considerado mais adequado foi então selecionado para dividir os objetos em comunidades.

2.3.4.4 Índices de validação

Tão importante quanto dividir um conjunto de dados em comunidades é determinar a qualidade dessa divisão. Esta análise poderia ser efetuada por um experimentador, no entanto, a subjetividade desse julgamento motivou a criação de índices de validação, medidas que buscam quantificar numericamente cada particionamento. Esses índices podem ser divididos em duas categorias: índices de validação internos, que avaliam as comunidades apenas tomando por referência informações intrínsecas aos dados, e índices de validação externos, que avaliam o resultado com base em informações prévias a respeito do conjunto analisado (RENDóN *et al.*, 2011; LIU *et al.*, 2013).

Tendo em vista o foco deste trabalho no processamento de trens de *spike* e o fato de carecermos de informações prévias sobre a organização funcional dos neurônios registrados, optamos pelo uso de índices de validação interna. De maneira geral, esses índices tendem a atribuir uma qualidade maior a comunidades bem separadas e compactas (GUERRA *et al.*, 2012). Os índices selecionados foram os de Davies-Bouldin e Silhouette, que serão descritos na próxima Seção.

2.3.4.4.1 Índice de Davies-Bouldin

Davies e Bouldin (1979) propõem um índice que é a razão da soma da dispersão no interior de uma comunidade C_k pela separação entre comunidades distintas C_k e C_l, como pode ser verificado pela Equação 2.35, onde N_c representa o número total de comunidades,

$\overline{d}_{C_k}$ é a distância média de cada ponto da comunidade C_k ao centróide e d_{C_k,C_l} é a distância euclidiana entre os centróides das comunidades C_k e C_l.

$$DB = \frac{1}{N_c}\sum_{k=1}^{N_c} max_{k\neq l}\frac{\overline{d}_{C_k} + \overline{d}_{C_l}}{d_{C_k,C_l}} \tag{2.35}$$

Podemos observar na expressão que a medida que o índice diminui, a distância entre comunidades distintas aumenta e a dispersão dos objetos da comunidade diminui. No entanto, a medida que o número de comunidades se aproxima do número de objetos, o índice tende a zero. Assim, para evitar a solução trivial, em que todos os objetos são independentes entre si, optamos pela partição correspondente ao primeiro ponto de inflexão do índice DB. Além disso, nossa diretriz foi obter o menor número possível de comunidades, correspondentes a agrupamentos maiores de células. Em um exemplo ilustrativo, no caso da Figura 2.5, isso significa que demos preferência ao cenário que identificava duas comunidades em detrimento do cenário que apontava seis comunidades. Para uma função contínua e diferenciável, define-se a ocorrência um ponto de inflexão quando há mudança na concavidade da função. No entanto, o particionamento em diferentes comunidades é uma função discreta, com domínio no conjunto de números inteiros. Em nosso problema, analisamos valores inteiros i 1 até $n_{Neurons}$. Para o índice $DB(i)$ calculado de maneira discreta, detectamos mudanças na concavidade através da mudança mudança de sinal de $DB(i+1) - DB(i)$. Assim, o conjunto foi dividido em N_c comunidades, onde N_c era o menor valor de i, tal que $DB(i+1) - DB(i) \geq \epsilon$, em que $\epsilon = 0$.

2.3.4.4.2 Índice Silhouette

O índice Silhouette (ROUSSEEUW, 1987) é calculado pela comparação das relações de um objeto X com os demais pertencentes a sua comunidade C_k e com os pertencentes a comunidades distintas, como pode ser verificado na Equação 2.36, a seguir. $a(X)$ corresponde à dissimilaridade média de X com todos os demais objetos de C_k, e pode ser interpretado como uma medida do quão adequado seria atribuir X à comunidade C_k. $b(X)$, por sua vez, recebe o valor referente à menor dissimilaridade de X com outras comunidades, ou seja, se $b(X)$ é elevado, isso significa que alocar X às comunidades vizinhas não seria adequado. Neste trabalho, o índice utilizou a distância euclidiana para o cálculo da dissimilaridade.

$$Sil(X) = \frac{b(X) - a(X)}{max(b(X), a(X))} \tag{2.36}$$

É possível observar que valores elevados de $Sil(X)$ correspondem à situação desejada, já que ocorrem quando $a(X) << b(X)$. Como o índice é computado localmente para X, calculamos então uma média para todo o conjunto de objetos ($\overline{Sil}$). No entanto, de maneira semelhante ao índice de Davies-Bouldin (Seção 2.3.4.4.1), a medida que o número de comunidades se aproxima do número de objetos, o índice $\overline{Sil}$ também converge para um valor limite, neste caso, 1. De maneira análoga, para evitarmos a solução trivial também optamos a partição correspondente ao primeiro ponto de inflexão de $\overline{Sil}$, ou seja, o número de comunidades N_c era tal que N_c recebia o menor valor de i tal que $\overline{Sil}(i+1) - \overline{Sil}(i) \leq \epsilon$, em que $\epsilon = 0$.

2.3.5 Estabilidade de uma partição

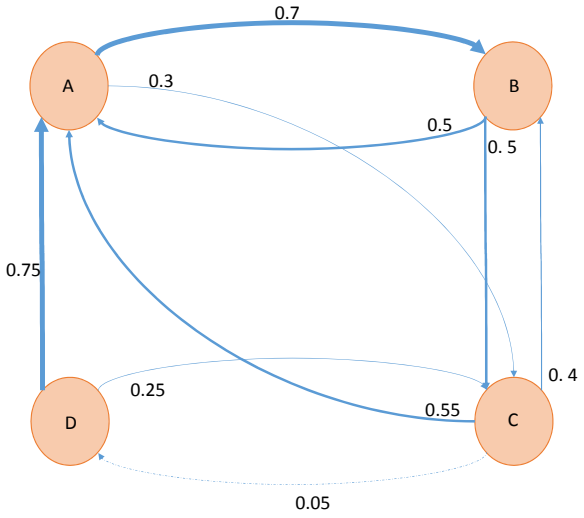
Este algoritmo também utiliza um modelo de grafos. No entanto, diferente da Modularidade, descrita na Seção 2.3.3, o método incorpora na noção de comunidades uma análise de sua dinâmica ao longo do tempo, como será descrito a seguir.

Ao proporem a metodologia, Delvenne *et al.* (2010) observaram que a dinâmica do grafo ao longo do tempo está bastante relacionada a sua topologia e é um fator determinante para a definição de comunidades. Afinal, em escalas temporais menores é esperada a ocorrência de múltiplos agrupamentos pequenos e de curta duração, enquanto que para maiores escalas temporais, haverá um número menor de comunidades, porém, com uma dinâmica persistente.

Como mencionado anteriormente na Seção 2.3.3, a existência de comunidades em um grafo está associada à uma maior conectividade em seu interior do que com seu exterior. Isso implica que, ao selecionarmos um nó e transicionarmos aleatoriamente para outro nó conectado a ele, teremos uma maior probabilidade de permanecermos na comunidade do que de sairmos dela. Essa seleção e movimentação aleatória recebe em inglês o nome de *random walk*. O algoritmo consiste em realizar *random walks* ao longo do grafo e identificar as comunidades por meio de observações das transições e das tendências de agrupamento em diferentes escalas temporais (BILLEH *et al.*, 2014).

Matematicamente, a análise da dinâmica começa ao estabelecermos um paralelo entre a estrutura da rede, modelada segundo um grafo, com um modelamento segundo uma cadeia de Markov. Uma cadeia de Markov consiste num modelamento de uma situação segundo um espaço de estados, em que a transição para o próximo estado é independente de ocorrências passadas. Um espaço de estados, por sua vez, pode ser entendido como o conjunto de condições que um objeto pode assumir e a probabilidade de ele transicionar de uma condição para outra. Uma representação esquemática para a matriz abaixo pode ser conferida na Figura 2.16. Observe que se iniciarmos uma *random walk* num nó aleatório e realizarmos transições

Figura 2.16 – Representação esquemática de uma matriz segundo um espaço de estados. A similaridade entre diferentes objetos passa a ser encarada como a probabilidade de transição de um nó para outro. Para facilitar a visualização, setas mais grossas correspondem a probabilidades de transição maiores. Note que a soma das probabilidades de transição é igual a 1. Os nós A,b e C formam uma partição estável ao longo do tempo de Markov, definindo uma comunidade.



Fonte: Produzida pela autora

consecutivas de um nó para outro tenderemos a permanecer nos nós A,B e C que, então, definirão uma partição estável e serão associados à formação de uma comunidade.

objetos	A	B	C	D
A	0	0.7	0.3	0
B	0.5	0	0.5	0
C	0.55	0.4	0	0.05
D	0.75	0	0.25	0

Naturalmente, o processo pode sofrer variações já que sua inicialização é aleatória, portanto, para elevar a robustez, ele deve ser repetido múltiplas vezes para cada valor da escala temporal da cadeia de Markov. Apesar de ser natural que esses agrupamentos variem com a escala temporal de Markov, é esperado que uma divisão da rede em comunidades permaneça estável ao longo de uma faixa de tempo. Essa estabilidade dinâmica de uma partição será adicionada ao conceito de comunidade segundo o ponto de vista do modelo de grafos. É importante observar que o tempo de Markov, correspondente a transições no espaço

de estados, é distinto do tempo físico (BILLEH *et al.*, 2014). Além disso, a estabilidade de uma partição realizada para uma única transição, ou seja, para o tempo de Markov igual a 1, corresponde à Modularidade (DELVENNE *et al.*, 2010).

Billeh *et al.* (2014) propuseram o uso deste algoritmo no contexto de detecção de ANs. Além de quantificarem a estabilidade das partições, eles também quantificaram a invariabilidade dos grupos por meio de um índice denominado Variação de Informação (VI), originalmente proposto por Meila (2007), que, se próximo a 0, indica um particionamento robusto e, não robusto, se próximo a 1. Assim, o experimentador selecionaria as ANs ao escolher situações estáveis com a escala temporal de Markov, associadas à uma variação de informação com poucas flutuações e em torno de zero.

Apesar de a partição de um conjunto de dados não possuir uma única solução, conforme já mencionado e ilustrado na Seção 2.3, neste trabalho padronizamos a partição selecionada referente ao maior *plateau* em que VI=0 ao longo do tempo de Markov variando de 10^{-2} até 10^{2} num passo de $10^{0.05}$.

2.3.6 Adaptações

Nesta seção descrevemos o processamento realizado pelos algoritmos de detecção de comunidades a partir da matriz de similaridade S calculada segundo alguma das metodologias descritas na Seção 2.2. Observamos que os algoritmos, da forma como originalmente implementados no contexto de detecção de ANs, recebiam como entrada matrizes de similaridade com relações distintas, de similaridade ou de distância, conforme descrito na Seção 2.2.4.

A análise de componentes principais recebe como entrada uma matriz de correlação, que se caracteriza por relações de similaridade entre os objetos, ou seja, as diagonais principais da matriz são iguais a 1. Dessa maneira, matrizes computadas com métricas distintas foram ajustadas segundo a Equação 2.23 para que a decomposição em valores singulares e a análise de componentes independentes também trabalhassem com esse tipo de relação.

Analogamente, as matrizes de similaridade utilizadas pela Modularidade, pela análise de agrupamento hierárquico e pela estabilidade de uma partição apresentavam relações de distância entre seus elementos. Dessa maneira, quando a matriz de similaridade recebida por esses algoritmos era a de correlação, efetuávamos a conversão para relações de distância segundo a Equação 2.23.

Outra adaptação implementada refere-se à detecção de comunidades. Notamos que, em algumas situações, os algoritmos de agrupamento hierárquico e o de estabilidade de uma partição caracterizavam um único objeto como compondo uma comunidade. Uma correção

foi implementada, de maneira que uma comunidade deveria conter ao menos dois objetos.

2.3.7 Comentário geral sobre as limitações dos métodos de classificação

Nesta seção revisamos alguns dos algoritmos empregados na detecção de comunidades e que já foram aplicados com sucesso em instâncias de nosso problema alvo: a identificação de ANs. A maioria desses algoritmos foi proposta no contexto de *data mining*, em que a noção de comunidade não é um conceito bem definido. Essa diversidade de algoritmos e conceitos é importante pois permite a exploração de diferentes relações entre os dados. No entanto, é necessário um olhar crítico do experimentador para que as diferentes propriedades e conceitos intrínsecos a cada algoritmo não levem a confusões e interpretações equivocadas (ESTIVILL-CASTRO, 2002) .

Esse olhar crítico deve também ser estendido na análise de ANs, a fim de que eventuais limitações dos algoritmos estejam mapeadas, que as hipóteses que definem uma comunidade sejam entendidas e, principalmente, que essas hipóteses sejam compatíveis com o comportamento de células reais.

É importante também considerar que a escolha do algoritmo deve também levar em conta sua complexidade e custo computacional e, principalmente, a aplicação à que os dados se destinam. Muitas vezes um algoritmo de maior complexidade pode apresentar um desempenho superior ou na detecção de ANs ou na determinação dos neurônios que as compõe, mas essa melhoria pode não ser benéfica o suficiente dado o tempo de processamento requerido. De maneira semelhante, cada aplicação possui características diferentes, que podem justificar a escolha de um algoritmo em detrimento de outro.

Este trabalho pretende abordar as questões levantadas nos parágrafos anteriores. Estudaremos os contextos de aplicação dos algoritmos, fazendo sempre que possível um paralelo com o comportamento de redes neuronais reais, e clarificaremos as limitações que cada um deles apresenta.

Na próxima seção detalharemos as métricas utilizadas na quantificação do desempenho do processamento aplicado.

2.4 Métricas de desempenho

Nesta seção faremos uma descrição da metodologia utilizada para quantificar, de forma numérica, o desempenho de um processamento. A comparação dos resultados nos permitirá comparar a performance de um algoritmo frente a outro. Dessa maneira, propusemos definições de métricas de desempenho que considerem as seguintes características de associação: 1)

número de assembleias detectadas (Seção 2.4.1) e 2) classificação dos neurônios (Seção 2.4.2). A escolha por duas métricas tem por objetivo evitar conclusões incorretas pois, por exemplo, um algoritmo pode ser excelente na detecção do número de assembleias, mas deficiente na detecção dos neurônios que pertencem àquela assembleia. Além disso, o processamento dos sinais pode ter objetivos diferentes dependendo da aplicação a se que destina, sendo em alguns casos mais interessante a correta identificação de grupos e, em outros, uma classificação automática de cada elemento da rede. Um experimentador pode, se desejar, combinar as métricas propostas em um único índice de acordo com a aplicação.

2.4.1 Número de assembleias detectadas

A métrica de desempenho para o número de assembleias detectadas ($P_{n_{Ass}}$) é calculada segundo a equação:

$$P_{n_{Ass}} = \begin{cases} 1 - \frac{|nda - nra|}{nra}, \ se \quad 0 < nda < 2nra \\ 0, caso\ contrário \end{cases} \tag{2.37}$$

onde:

- nda é o número de assembleias detectadas pelo algoritmo.
- nra é o número de assembleias reais, pré-definidas na geração do dado simulado.

Essa métrica assume um desempenho decaindo linearmente com respeito ao número de assembleias detectadas erroneamente ou não detectadas. Em outras palavras, o efeito de um falso positivo e de um falso negativo é equalizado.

2.4.2 Classificação dos neurônios

Estabelecemos um método para quantificar a similaridade entre os neurônios das assembleias identificadas e os de cada assembleia pré-definida. Ele é feito ao computarmos a denominada "Matriz de Comparação", ($C_{ndaXnra}$).

Cada elemento $c_{i,j}$ de C é calculado segundo a expressão:

$$\begin{cases} c(i,j) = \begin{cases} \frac{diff(i,j)}{nReal(i)}, \quad se \quad diff(i,j) > 0 \\ 0, caso\ contrário \end{cases} \\ diff(i,j) = N_E(i,j) - N_D(i,j) \end{cases} \tag{2.38}$$

onde:

- $N_E(i,j)$ corresponde ao número de neurônios que pertencem tanto à assembleia especificada i quanto à detectada j.
- $N_D(i,j)$ corresponde ao número de neurônios que não estão de acordo entre a assembleia estabelecida i e à identificada j (falso positivos).
- $nReal(i)$ é o número de neurônios pré-definidos na assembleis especificada$i$.

Ao analisarmos a matriz C, podemos determinar quais das ANs encontradas são mais similares às especificadas. Portanto, o cálculo da métrica que computa a classificação dos neurônios P_{mem} pode ser efetuada segundo a expressão a seguir:

$$P_{mem} = \sum_{k=1}^{nda} \frac{(C(l,k)|_{l=1}^{nra})}{nra} \tag{2.39}$$

Essa métrica assume que um neurônio falso positivo irá penalizar o índice de maneira a cancelar a contribuição de outro que tenha sido corretamente identificado. Embora a Equação 2.39 aparentemente capture o efeito do número total de assembleias, falta a ela penalizar ANs falsas positivas, sendo, portanto, complementada pela Equação 2.37.

2.4.3 Exemplo

Nessa seção forneceremos um exemplo que ilustrará o efeito que cada uma das métricas propostas neste trabalho captura.

Voltemos ao exemplo ilustrado nas Figuras 2.8a, 2.8c e 2.8e. Nessa situação, a simulação de 20 células contava com três ANs A_i, i=1,2,3, compostas pelos seguintes elementos: $A_1 = \{\#1, \#2, \#3, \#4\}$; $A_2 = \{\#4, \#5, \#6, \#7\}$ e $A_3 = \{\#8, \#9, \#10\}$. As comunidades encontradas pelo algoritmo A'_i foram: $A'_1 = \{\#1, \#2, \#3, \#4, \#6, \#7\}$, $A'_2 = \{\#1, \#2, \#3, \#5, \#6, \#7\}$ e $A'_3 = \{\#8, \#9, \#10\}$, o que implica em $P_{mem} = 0.5$, e, $P_{n_{Ass}} = 1$, ou seja, os neurônios de cada AN não foram corretamente identificados ainda que o número de ANs encontradas seja numericamente igual ao inserido nos dados artificiais. Dessa maneira, se uma determinada aplicação estiver interessada apenas na detecção de diferentes ANs, a métrica da Equação 2.37 nos mostrará que o PCA é adequado para essa situação. No entanto, caso se deseje acompanhar a atividade dos membros das ANs, devemos atentar para o resultado da métrica da Equação 2.39.

Na próxima seção, concluimos este capítulo descrevendo a tarefa comportamental que será analisada nesse trabalho, além de apresentar a hipótese que pretendemos validar nessa dissertação.

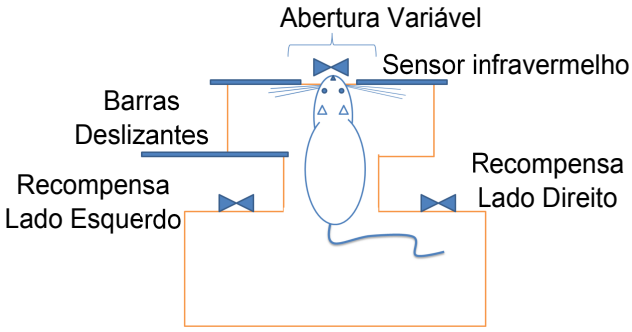
2.5 Descrição do experimento em modelo animal

Na Seção 1.1, apresentamos a modelagem computacional dos dados para o estudo de ANs no âmbito deste projeto. Como conclusão do trabalho, propomos a seguir um estudo prático da formação dinâmica de ANs, levando-se em consideração todos os aspectos relativos as métricas de similaridade e métodos de classificação apresentados.

O sistema de vibrissas de ratos é utilizado como um dos modelos para o estudo do sistema somatossensorial de mamíferos e é interessante do ponto de vista fisiológico devido à representação somatotópica ao longo de toda a via trigeminal. Seguindo a proposta de Krupa *et al.* (2001), utilizamos uma tarefa em que ratos *Long-Evans* foram treinados em uma caixa comportamental escura para, empregando apenas suas vibrissas, discriminar entre aberturas largas e estreitas. A caixa possuía alguns feixes de luz infravermelha, que, ao serem quebrados, marcam a passagem do animal pela barra de discriminação. A caixa era composta por duas câmaras, a de estimulação, onde ocorre a estimulação tátil, e a anterior, onde é fornecida a recompensa. A câmara de estimulação continha barras móveis que definiam, para cada repetição do experimento, aberturas largas ou estreitas. Para garantir a estimulação de todas as vibrissas, o animal deveria entrar na câmara, mover-se e colocar o focinho num orifício, o que quebrava um feixe luminoso e ativava a liberação de uma recompensa líquida na câmara anterior. Caso a abertura fosse larga, uma porta se abria liberando o bico de uma mangueira com água do lado direito, e, caso estreita, do lado esquerdo. Essas portas também contavam com sensores infravermelho e permitiam a medição do desempenho do animal, com base na taxa de acertos da correspondência dos lados com as aberturas. Uma imagem ilustrativa pode ser visualizada na Figura 2.17, abaixo.

Essa tarefa é interessante porque várias áreas distintas estão envolvidas no processamento da discriminação tátil. Krupa *et al.* (2001) encontraram evidências que a via trigeminal do sistema somatossensorial codifica as representações dos estímulos táteis envolvidos nessa tarefa. A seguir, Krupa *et al.* (2004) demonstraram a ocorrência de atividade antecipatória no córtex somatossensorial primário (S1). Além disso, Pais-Vieira *et al.* (2013b) mostraram a dependência da modulação antecipatória em S1 por neurônios do cótex motor primário e observaram a ocorrência de atividade antecipatória também em camadas (*layers*) específicas das vias lemniscal e paralemniscal do sistema trigeminal, como os núcleo talâmicos ventroposterior medial e posterior medial. Além disso, Zangaladze *et al.* (1999) utilizaram

Figura 2.17 – Desenho esquemático do aparato comportamental. O animal distinguia entre aberturas largas e estreitas utilizando suas vibrissas e associava cada cenário à obtenção de uma recompensa em um lado da câmara. Caso a abertura fosse larga, a recompensa líquida seria liberada no lado direito, se estreita, no lado esquerdo.



Fonte: Produzida pela autora

a estimulação magnética transcraniana (TMS) para perturbar a atividade no córtex visual primário (V1) e observaram interferência na discriminação tátil utilizada para a orientação espacial destes animais. Vasconcelos *et al.* (2011) registraram S1 e V1 simultaneamente, encontrando uma quantidade de informação similar nas respostas das duas áreas. Já Sieben *et al.* (2013) notaram uma relação entre as oscilações em S1 e estímulos visuais, fortalecendo a hipótese de que o processamento da informação em áreas corticais primárias é distribuído.

Também existem experimentos ressaltando o papel dos córtices parietal posterior (PPC) e pré-frontal (PFC) no processamento da informação tátil, pois, verificou-se anatomicamente que essas áreas recebem sinais de uma série de regiões sensoriais (REEP *et al.*, 1994; MILLER; COHEN, 2001; BEHRMANN *et al.*, 2004; ROMANSKI, 2011). Acredita-se que o PPC relacione-se a processos cognitivos de atenção (BEHRMANN *et al.*, 2004; STILLA *et al.*, 2007) e existem evidencias que ele está diretamente envolvido no processamento da representação espacial do ambiente (WHITLOCK *et al.*, 2008) e na associação entre informações táteis e visuais (WINTERS; REID, 2010). Já o PFC é associado a planejamento e raciocínio (MILLER; COHEN, 2001) e alguns estudos mostram que a área também participa da codificação de informações somatossensoriais (ROMANSKI, 2011).

Portanto, nossa hipótese é a de que os neurônios dessas áreas apresentarão ANs em instantes próximos aos de discriminação tátil.

Quando o animal atingia uma taxa de acerto de 75%, ele era implantado com microeletrodos para o registro de sua atividade eletrofisiológica extracelular. Para a construção das matrizes de microeletrodos foram utilizados fios de tungstênio de 50 μm revestidos com Teflon (California Fine Wire Company). As matrizes de microeletrodos foram construídas

para implante em 4 áreas corticais, sendo elas: PFC, S1-camadas 4 e 5, PPC e V1. As configurações de cada matriz de microeletrodos foram estabelecidas a partir das coordenadas estereotáxicas do atlas de Paxinos e Watson (2007), são elas:

- PFC
  AP: +2.20 a -0.40 mm, ML: 1.40 mm, DV: 1.96 mm, θ: 24 graus;
- S1
  AP: -2.04 a -3.24 mm, ML: 5.2 a 6.4 mm, DV: 800 μm (camada 4), DV: 1.2 mm (camada 5), θ: 39 graus;
- PPC
  AP: -3.24 a - 4.08 mm, ML: 1.8 a 3.3 mm, DV: 1.0 mm, θ: 90 graus;
- V1
  AP: -5.2 a -6.40 mm, ML: 3.8 a 5.0 mm, DV: 1.67 mm, θ: 20 graus.

Em cada implante neural foram utilizados 64 microeletrodos distribuídos nas 4 áreas, dispostos em matrizes na seguinte disposição:

- PFC: $1X8$ (espaçamento 200 μm), bilateral;
- S1: $4X4$ (espaçamento 400 μm), unilateral;
- PPC: $1X4$ + $2X6$ (espaçamento 300 μm), unilateral;
- V1: $4X4$ (espaçamento 400 μm), unilateral.

Todos os procedimentos foram aprovados pelo Comitê de Ética no Uso de Animais (CEUA) do IIN-ELS sob o protocolo 01/2013.

Este capítulo descreveu a modelagem computacional e experimental deste projeto, bem como as métricas de similaridade e métodos de classificação que serão avaliados quanto ao seu desempenho na detecção de ANs. No próximo capitulo, apresentamos os resultados e, no seguinte, as discussões.

3 Resultados

O objetivo deste trabalho é avaliar comparativamente o desempenho de diversas combinações de métricas de similaridade e métodos de classificação para a caracterização de ANs. Assim, inicialmente, exibiremos rodadas completas de análise, em que comparamos diferentes associações entre métricas de similaridade e algoritmos aplicados ao um mesmo conjunto de dados simulados. Consideramos neste estudo diferentes topologias de ANs, como exibido na Figura 2.1, que apresentam não só interações de excitação, que são as tradicionalmente tratadas, mas também inibições e atrasos temporais. Além disso, construímos conjuntos de dados que levam em conta interações mais complexas entre os trens (Seção 2.1.3), e comparamos o resultado com o modelo simplificado (Seção 2.1.2).

A seguir, apresentaremos os resultados de aplicarmos algumas metodologias em um conjunto de registros eletrofisiológicos extracelulares de ratos *Long Evans* realizando uma tarefa de discriminação tátil (Seção 2.5).

3.1 Simulações

Para produzir dados artificiais com propriedades de dados biológicos reais, utilizamos como referência os registros extracelulares de um rato *Long-Evans* realizando a tarefa comportamental de discriminação tátil descrita na Seção 2.5. Esse animal foi implantado nas áreas PFC, PPC e S1. Após o *spike sorting*, detectou-se um total de 42 unidades que foram gravadas em uma seção de 39 minutos com taxa de amostragem de $40000Hz$. A atividade da população neuronal resultou numa taxa média de disparos de 22,37 *spikes/segundo*.

3.1.1 Descrição das ANs simuladas

Conforme já mencionado na Seção 2.1.1 e exemplificado na Figura 2.1, nós analisamos três categorias de assembleias definidas com base na interseção de seus membros. Assim, alguns neurônios da população simulada foram selecionados para participar de nossas ANs artificiais A_i, i=1,2. No primeiro caso, simulamos uma população de 30 células (caso 1), e, no segundo (caso 2), de 60 células segundo as definições a seguir. Todas as outras células permaneceram com uma atividade independente. Nossos trens artificiais foram criados com um comprimento de 10^4 ms.

Caso 1:

- Assembleias disjuntas: A_1={#1,...,#10}; A_2={#11,...,#20}.
- Assembleias com pelo menos um neurônio exclusivo: A_1={#1,...,#10}; A_2={#6,...,#15}.
- Assembleias sem neurônios exclusivos: A_1={#1,...,#10}; A_2={#6,...,#10}.

Caso 2:

- Assembleias disjuntas: A_1={#1,...,#25}; A_2={#26,...,#50}.
- Assembleias com pelo menos um neurônio exclusivo: A_1={#1,...,#25}; A_2={#21,...,#45}.
- Assembleias sem neurônios exclusivos: A_1={#1,...,#25}; A_2={#16,...,#25}.

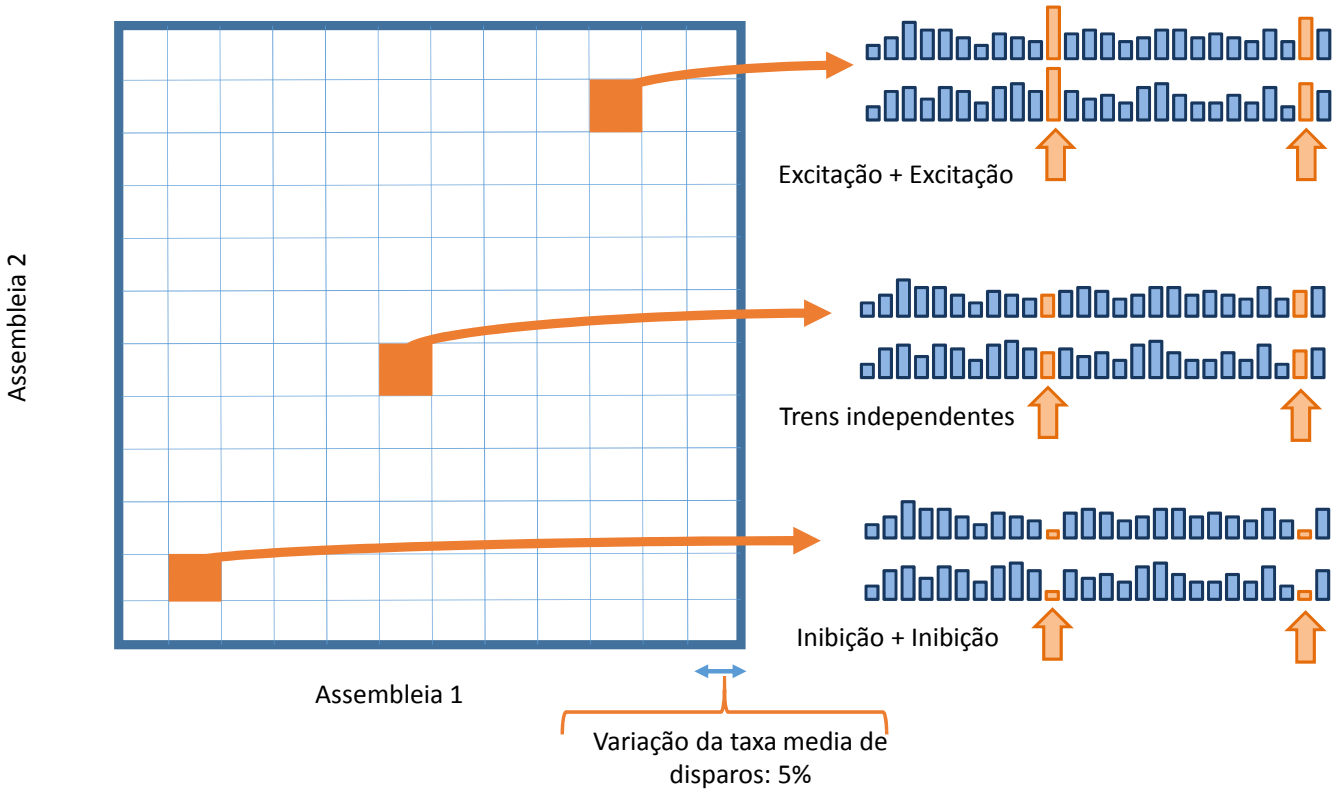
Para os trens de *spike* gerados segundo a distribuição de Poisson, nós estabelecemos interações entre as células que pertenciam às ANs em 300 *bins* coincidentes e selecionados aleatoriamente. Cada *bin* possuia uma duração de 10 ms. Neste estudo não avaliamos o efeito do aumento ou diminuição do número de coincidências no resultado.

Nós analisamos a variação do desempenho ao modificarmos gradualmente a taxa média de disparos desses *bins* aleatoriamente selecionados, num passo de 5%, para a faixa de -95% até 95% em relação ao valor basal. A seguir, a análise foi repetida para os casos 1 e 2 (descritos acima) com assembleias disjuntas, mas, dessa vez, inserimos o atraso de um *bin* para os cinco últimos neurônios de A_1.

Os resultados foram exibidos seguindo a disposição esquematizada pela Figura 3.1, a seguir. Cada quadrado demarcado no gráfico exibiu o resultado do método aplicado a uma população neuronal simulada, em que as interações entre os membros das ANs variavam ao longo dos eixos. Valores negativos simulavam uma queda local da taxa média de disparos, o que representava interações de inibição. Na região central, as variações eram muito baixas e, portanto, pouco diferiam de um cenário em que os trens eram independentes. Finalmente, os valores positivos representavam aumentos na taxa de disparo, e simulavam interações de excitação. Como os quadrados possuem o mesmo tamanho, omitimos as linhas nas figuras que exibem o desempenho, para maior clareza.

Exibimos de maneira análoga os resultados para as simulações de trens de *spike* modelados por interações de segunda ordem, seguindo a metodologia descrita na Seção 2.1.3. Neste caso, interações inibitórias e excitatórias não foram mais descritas somente por variações momentâneas na taxa média de disparos, mas pela correlação dos trens de *spike* em

Figura 3.1 – Representação esquemática utilizada para a disposição dos resultados das simulações. Avaliamos o desempenho de cada forma de processamento cobrindo uma grande variedade de cenários, por meio de rodadas extensivas de simulação. As interações entre os neurônios que compõem as AN foram descritas por variações locais da taxa média de disparos, que podem ser visualizadas ao longo dos eixos da Figura. Inibições foram caracterizadas por valores negativos e excitações, por valores positivos.



Fonte: Produzida pela autora

bins. Neste trabalho, utilizamos *bins* de 10 *ms*. Para simplificar o número de combinações possíveis, definimos que a correlação cruzada entre os elementos de uma mesma AN assumiriam valores iguais e, no caso de um neurônio participar de mais de uma assembleia, a correlação resultante seria computada pela soma dos valores associados a cada AN. As linhas pontilhadas pretas foram mantidas nas figuras para explicitar que os pontos não eram igualmente distribuídos, tendo em vista que foram extraídos de dados reais. Assim, para inferir conclusões gerais, nesses casos, restringimos nossa análise à região central, já que ela continha uma maior concentração de pontos, e excluímos as bordas.

3.1.2 Resultados: ICA

Nesta seção exibimos o desempenho do algoritmo ICA à medida que alternamos a maneira de quantificar a similaridade entre os trens. O objetivo foi, a partir do mesmo conjunto de dados, fazer uma análise comparativa das métricas de similaridade. Começamos exibindo os resultados para ANs disjuntas (Seção 3.1.2.1), seguindo para ANs com pelo menos um neurônio exclusivo (Seção 3.1.2.2) e finalizamos com ANs sem neurônios exclusivos (Seção 3.1.2.3).

3.1.2.1 ICA: Simulações de ANs independentes

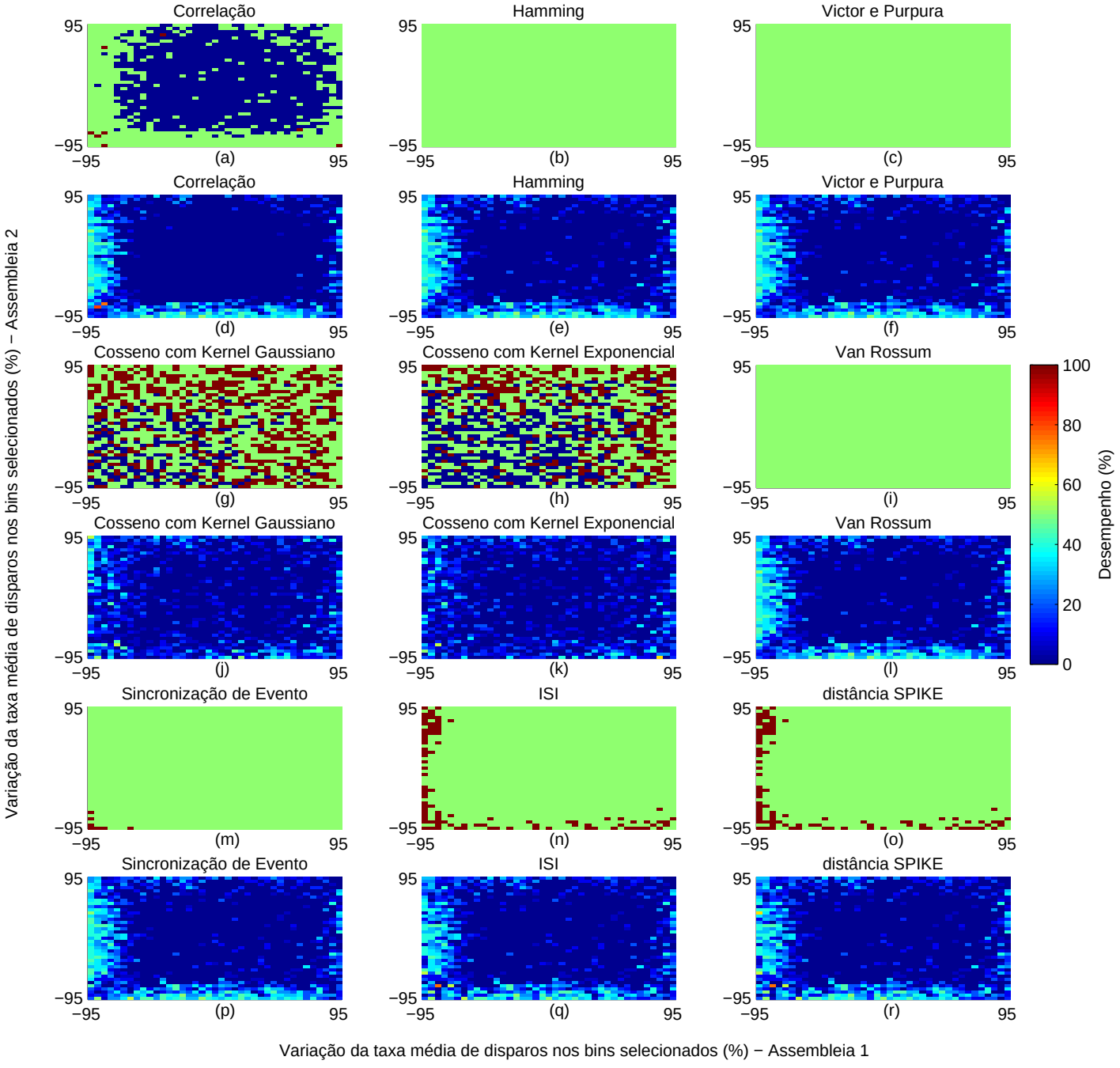
Para a análise da resposta do ICA aplicado a ANs independentes, começamos com trens simulados segundo a distribuição de Poisson e, a seguir, passamos para simulações que levam em conta também a covariância cruzada entre os neurônios.

3.1.2.1.1 ICA: Simulações utilizando trens de *spike* segundo a distribuição de Poisson em ANs independentes

Inicialmente, utilizamos trens artificiais gerados segundo o modelo de Poisson para um total de 30 células (caso 1, descrito na Seção 3.1.1) e ANs disjuntas. A Figuras 3.2 exibe os resultados de aplicação do ICA às métricas de desempenho relativas ao número de assembleias detectadas (linhas 1, 3 e 5) e à classificação dos neurônios (linhas 2, 4 e 6). Foi possível observar pela a análise do desempenho segundo o número de ANs detectadas (Equação 2.37) que o ICA não produziu resultados verossímeis quando associado a métricas de similaridade alternativas à correlação pois o algoritmo apontou a existência de ANs mesmo em trens de *spike* independentes (regiões centrais das figuras). Além disso, notamos que, para este caso, a classificação dos neurônios (linhas 2, 4 e 6 da Figura 3.2, Equação 2.39) foi similar para todas as métricas de similaridade e que mesmo métricas capazes de quantificar padrões de disparos mais complexos entre os neurônios não apresentaram melhorias significativas de desempenho.

Analisamos também o caso de uma população com as mesmas características, porém, com uma taxa média de disparos maior, de 64 *spikes/segundo* pois, nesta situação, notamos que as diferenças entre as métricas de similaridade foram mais expressivas (Figura 3.3, onde as linhas ímpares estão associadas à métricas de desempenho que quantifica o número de ANs detectadas, Equação 2.37 e, as pares, à classificação dos neurônios, Equação 2.39). Ressaltamos, porém, que este valor não é realista e foi utilizado apenas para extrapolarmos o desempenho do algoritmo com o aumento da taxa média de disparos. Notamos uma melhoria na detecção de comunidades para a Correlação (Figura 3.3a) quando as taxas de disparo

Figura 3.2 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1, linhas 1, 3 e 5) e na classificação dos neurônios (Seção 2.4.2, linhas 2, 4 e 6) quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 30 células e ANs independentes, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.



tinham uma maior variação, ou seja, nos cantos da Figura. Uma melhoria similar, porém, concentrada em regiões de inibição também foi observada para as métricas de similaridade baseadas nas relações temporais dos disparos (Figuras 3.3m, 3.3n e 3.3o). As regiões em que o algoritmo acusou a presença de ANs de maneira equivocada, mesmo em trens independentes, persistiram. Para a classificação dos neurônios (linhas pares da Figura 3.3), observamos que as métricas de similaridade Correlação, Sincronização do Evento, ISI e distância SPIKE (Figuras 3.3d, 3.3p, 3.3q, 3.3r) apresentaram também um melhor desempenho comparativamente às demais em regiões de máxima variação da taxa média de disparos (cantos das figuras), mas, de maneira geral, o padrão observado para o desempenho de todas as métricas na classificação dos neurônios foi parecido.

A seguir, voltamos nossa análise para uma população com a taxa de disparos de 22,37 *spikes/segundo*, porém, desta vez, composta por 60 células com ANs maiores, mas ainda disjuntas (Caso 2, descrito na Seção 3.1.1). O resultado pode ser observado na Figura 3.4, a seguir, referentes à detecção do número de ANs (linhas ímpares), Equação 2.37, e à classificação das células (linhas pares), Equação 2.39, respectivamente.

Pudemos perceber na comparação com a Figura 3.2 que o algoritmo apresentou uma melhora de desempenho para a detecção de ANs tanto para a correlação (Figura 3.4a) quanto para as métricas de similaridade baseadas na relação temporal entre os *spikes* (Figuras 3.4m, 3.4n e 3.4o). Houve um aumento na região com desempenho máximo, mas, persistiram as inconsistências em regiões com trens independentes. Para o uso do cosseno como métrica de similaridade, observou-se que o número de ANs espuriamente detectadas nas regiões inibitórias aumentou, o que, consequentemente, diminuiu o desempenho nessa região nas Figuras 3.4g e 3.4h.

Para a classificação das células (linhas pares da Figura 3.4), notamos uma ligeira melhoria em relação às linhas pares da Figura 3.2 nas bordas dos gráficos, mas, nessas regiões, o desempenho permaneceu em torno de 50%.

A seguir, analisamos novamente o efeito de aumentarmos a taxa média de disparos da população para 64 *spikes/segundo*, mas, dessa vez, para um conjunto maior de células. Os resultados para o desempenho na detecção do número de ANs (Equação 2.37) e da classificação das células (Equação 2.39) podem ser verificados, respectivamente, nas linhas ímpares e pares da Figura 3.5. Foi possível observar o surgimento de regiões em que as duas ANs foram detectadas para as métricas de Victor e Purpura (Figura 3.5c) e de Van Rossum (Figura 3.5i). Estas regiões consistiam em iterações fortes de excitação (canto superior direito da figura) e inibição (canto inferior esquerdo da figura). Além disso, assim como na análise comparativa da simulação de uma população de 30 células, observamos melhorias para as métricas

Figura 3.3 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1, linhas 1, 3 e 5) e na classificação dos neurônios (Seção 2.4.2, linhas 2,3 e 6) quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 64 *spikes*/*segundo*. Simulamos uma população de 30 células e ANs independentes, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.

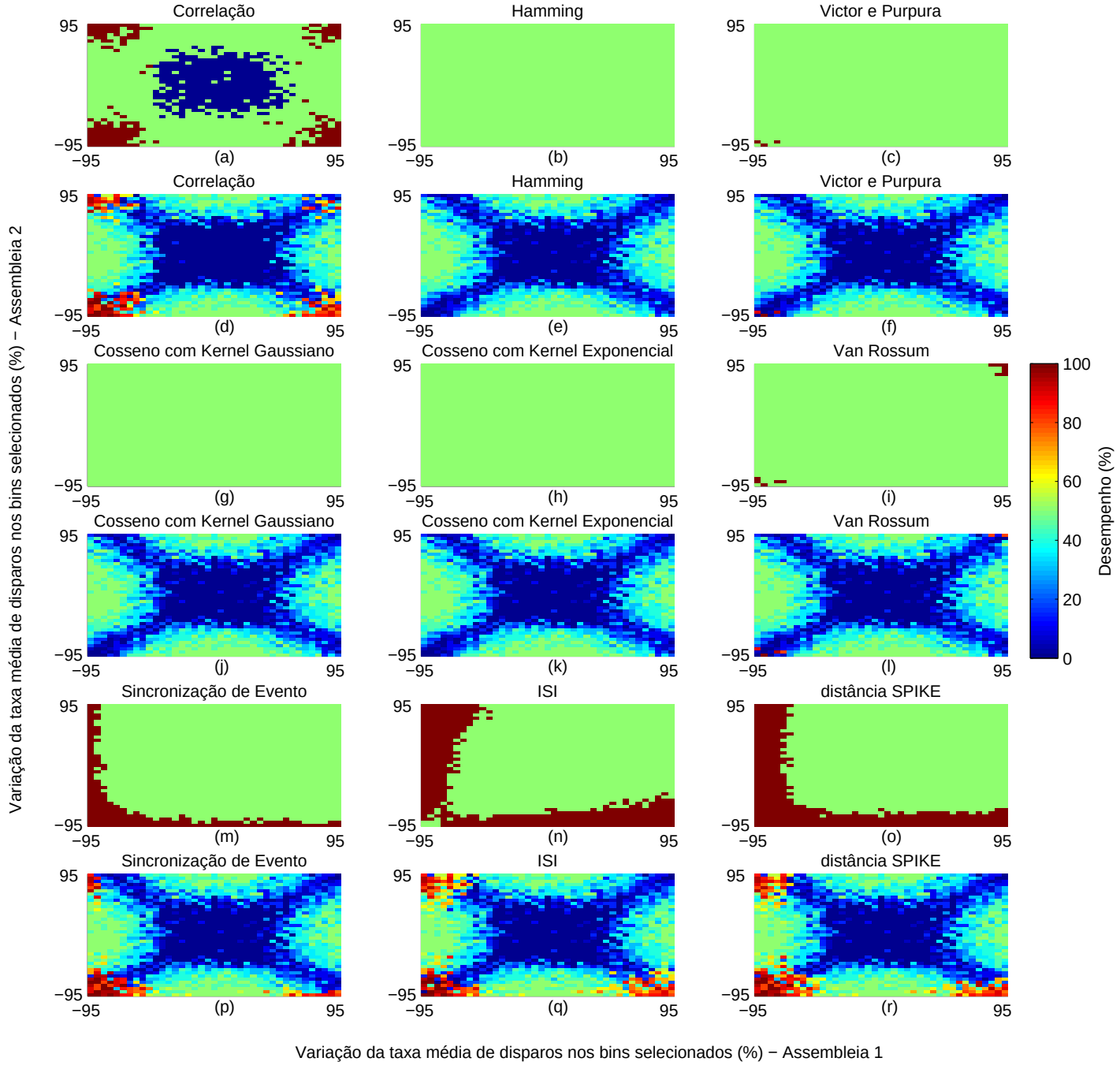
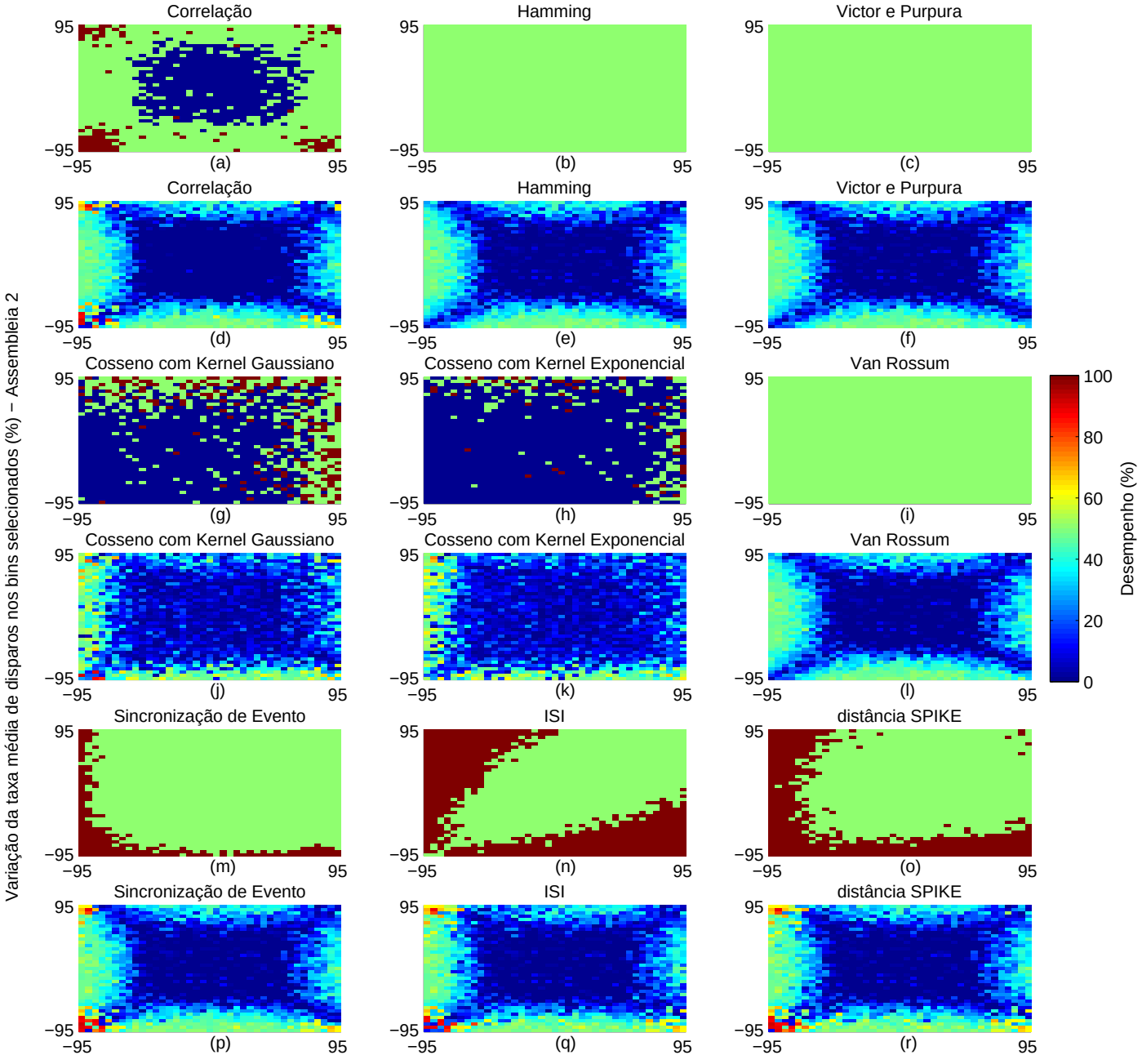


Figura 3.4 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1, linhas 1, 3 e 5) e na classificação dos neurônios (Seção 2.4.2, linhas 2,3 e 6) quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.



Correlação, Sincronização de evento, ISI e distância SPIKE com o aumento da taxa média de disparos tanto na detecção das ANs (Figuras 3.5a, 3.5m, 3.5n e 3.5o, respectivamente) quanto na classificação das células (Figuras 3.5d, 3.5p, 3.5q e 3.5r, respectivamente).

Na próxima etapa, avaliamos o efeito que a inserção de atrasos de fase no desempenho do algoritmo quando associado às diferentes métricas de similaridade. Para ressaltar os efeitos nas métricas de similaridade quando associadas ao ICA, optamos por exibir as simulações relativas à população de 60 células com taxa média de disparos de 64 *spikes/segundo*, já que, nestes casos, as diferenças entre as associações foram mais expressivas. O resultado para a detecção de ANs (Equação 2.37) segue nas linhas ímpares da Figura 3.6 e nas pares para a classificação das células (Equação 2.39), na Figura 3.6. Observamos uma queda no desempenho para a classificação das células, ao compararmos as linhas pares da Figuras 3.5 e 3.6, o que evidenciou que estas métricas de similaridade, ao serem associadas ao ICA, não foram robustas a atrasos temporais na interação entre as células membro de ANs. Além disso, tanto a correlação quanto a distância SPIKE mostraram diferenças também no desempenho para a identificação das ANs com a inserção de atrasos (Figuras 3.5a, 3.6a; e 3.5o, 3.6o), o que não foi observado para a Sincronização do evento e (Figuras 3.5m e 3.6m) para o ISI (Figuras 3.5n e 3.6n).

3.1.2.1.2 ICA: Simulações utilizando trens de *spike* que consideram a correlação entre neurônios em ANs independentes

A seguir, utilizamos em nossas simulações, além da taxa média de disparos, os valores de correlação cruzada dos trens de *spike* para definir as interações entre as ANs. Simulamos a atividade de uma população de 30 células, o caso 1 descrito na Seção 3.1.1, utilizando valores de correlação calculados para neurônios de um animal envolvido na tarefa de discriminação tátil. No entanto, este modelo produz um conjunto de dados que já descreve a atividade dos neurônios em taxas de disparos, e, portanto, no contexto deste trabalho, pode ser abordado apenas pelas métricas de similaridade da Seção 2.2.1. Os resultados podem ser conferidos nas Figuras 3.7a e 3.7b, referentes à aplicação do ICA às métricas de desempenho relativas ao número de assembleias detectadas (Equação 2.37); e nas Figuras 3.7c e 3.7d, à classificação dos neurônios (Equação 2.39).

Observamos que, para o conjunto de dados utilizados na escala temporal de 10 ms, houve um pequeno número de ocorrências de correlações negativas maiores que 0.005 em módulo, o que não possibilitou a caracterização de ANs inibitórias. Ao compararmos os resultados deste modelo na região que representa ANs excitatórias na Figura 3.7 com os obtidos pelo modelo simplificado de Poisson, na Figura 3.2a, 3.2d, 3.2b e 3.2e notamos que a métrica de Hamming retornou um desempenho similar, com uma queda de desempenho em

Figura 3.5 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1, linhas 1, 3 e 5) e na classificação dos neurônios (Seção 2.4.2, linhas 2,3 e 6) quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.

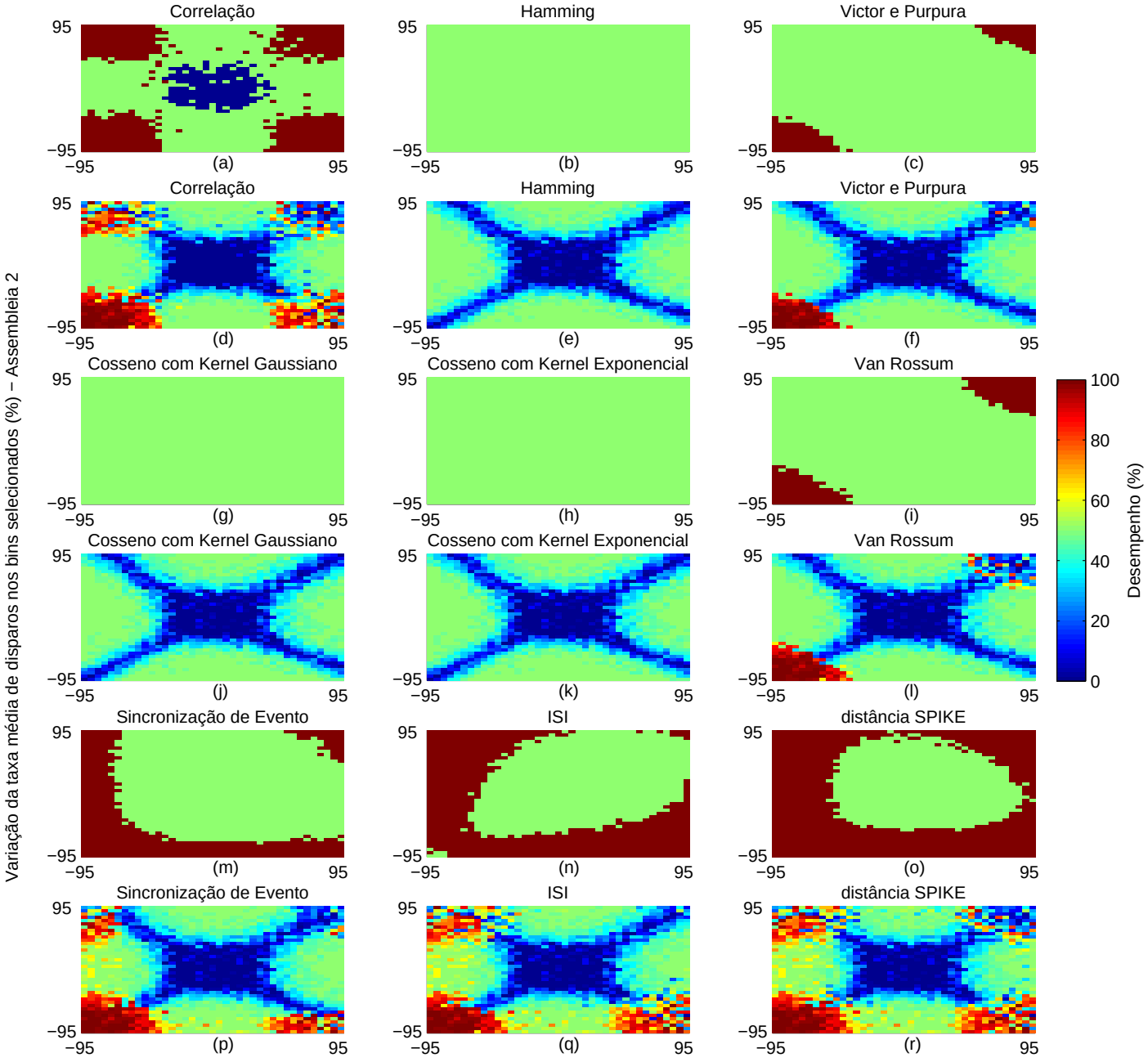
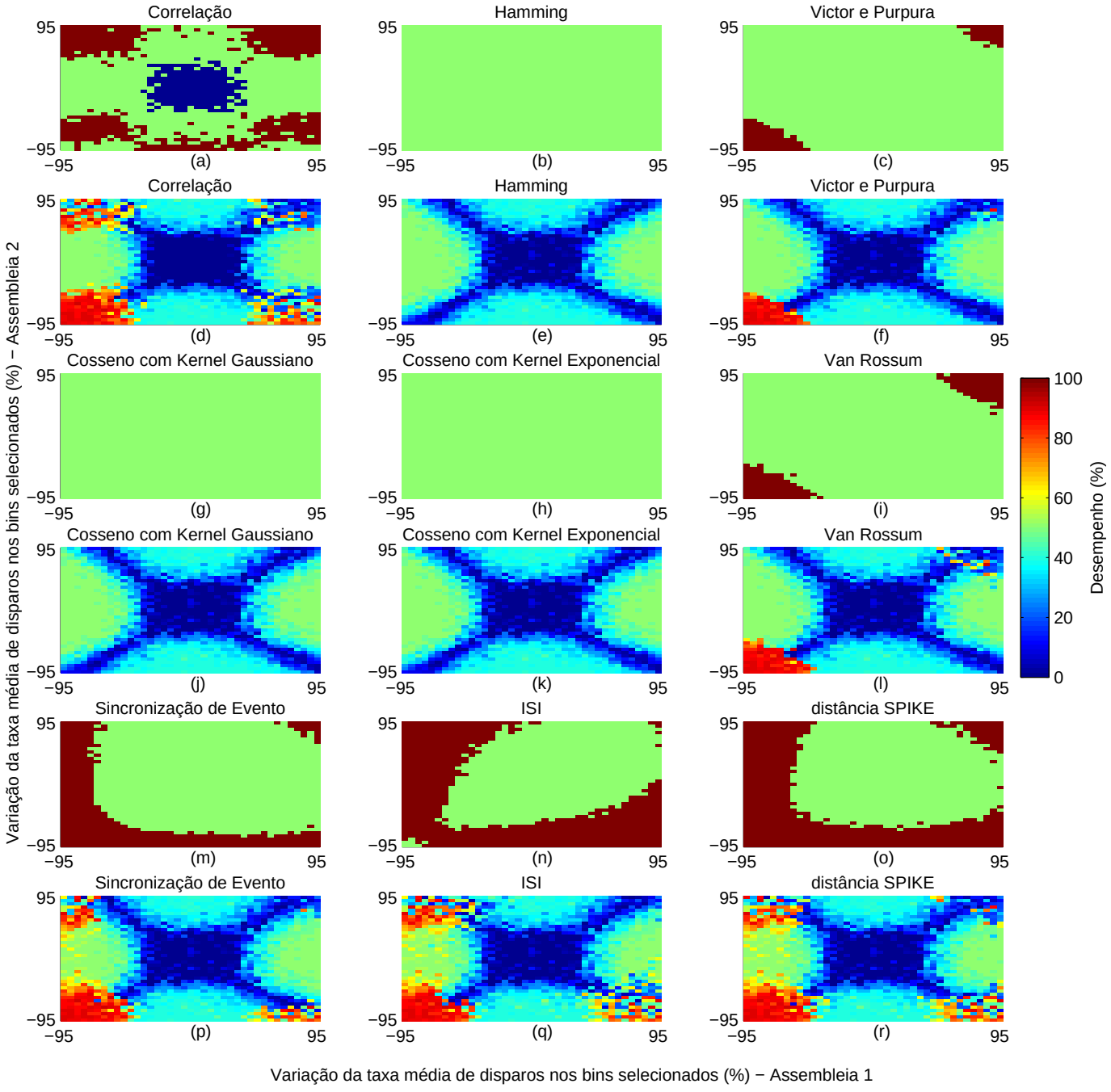


Figura 3.6 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1, linhas 1, 3 e 5) e na classificação dos neurônios (Seção 2.4.2, linhas 2,3 e 6) quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. A atividade de 5 neurônios de uma das ANs foi atrasada um *bin*.



uma pequena região para a classificação dos neurônios, enquanto que, a correlação apresentou uma melhoria significativa tanto para a detecção das ANs quanto para a classificação dos neurônios.

3.1.2.2 ICA: Simulações de ANs com ao menos um neurônio exclusivo

Nesta próxima etapa simulamos a resposta do ICA aplicado a ANs que compartilhavam neurônios (casos 1 e 2, descritos na Seção 3.1.1). Assim como na Seção anterior, iniciamos com a resposta do algoritmo aplicado a trens simulados utilizando a distribuição de Poisson (Seção 3.1.2.2.1) e, a seguir, partimos para o modelo com a correlação cruzada entre as células (Seção 3.1.2.2.2).

3.1.2.2.1 ICA: Simulações utilizando trens de *spike* segundo a distribuição de Poisson em ANs com ao menos um neurônio exclusivo

O desempenho da aplicação do ICA para os conjuntos de dados artificiais simulados utilizando o modelo de Poisson foi, em muitos casos, bastante similar para o exibido para ANs disjuntas. Portanto, optamos por exibir apenas os casos em que as diferenças foram expressivas.

Para a simulação de uma população de 30 células (caso 1, descrito na Seção 3.1.1) utilizamos inicialmente uma taxa média de disparos igual a 22,37 *spikes/segundo*, numericamente igual à obtida para um de nossos conjuntos de dados em que o animal realizava a tarefa de discriminação tátil descrita na Seção 2.5. Os resultados podem ser verificados na Figura 3.8, a seguir, para as métricas de correlação, ISI e distância SPIKE. Na primeira linha, os diagramas exibem os resultados correspondentes às métricas de desempenho relativas ao número de assembleias detectadas (Equação 2.37) e, na segunda, à classificação dos neurônios (Equação 2.39).

Observamos que, em comparação à Figura 3.2, há uma queda de desempenho na detecção de ANs para a correlação quando a interação entre as células é de excitação (primeiro quadrante da Figura 3.8a), acompanhada por uma queda na mesma região para a detecção de neurônios (Figura 3.8d). Além disso, o uso das métricas ISI e distância SPIKE registraram um aumento no desempenho para a detecção de comunidades à medida que as interações de inibição eram mais fortes (terceiro quadrante das Figuras 3.8b e 3.8c). Este aumento não foi acompanhado pela detecção de neurônios (terceiro quadrante das Figuras 3.8e e 3.8f).

A seguir, analisamos o efeito que o aumento da taxa de disparos teria sobre esta mesma topologia, portanto, computamos uma nova simulação com uma taxa média de disparos de 64 *spikes/segundo*. Os resultados podem ser observados na Figura 3.9. Na primeira

Figura 3.7 – Desempenho do ICA associado a diferentes métricas de similaridade quando aplicado a trens artificiais que modelam ANs independentes gerados segundo o modelo de correlação entre os neurônios (MACKE *et al.*, 2009) com uma taxa de disparos basal de 22,37 *spikes/segundo*. Na primeira linha, exibimos o desempenho para o número de assembleias detectadas (Equação 2.37)e, na segunda, para a classificação dos neurônios (Equação 2.4.2). As linhas pontilhadas pretas em cada um dos eixos delimitam os 39 valores distintos de correlação cruzada (extraídos de dados reais) e foram inseridas para salientar que os valores de correlação não eram uniformemente distribuídos.

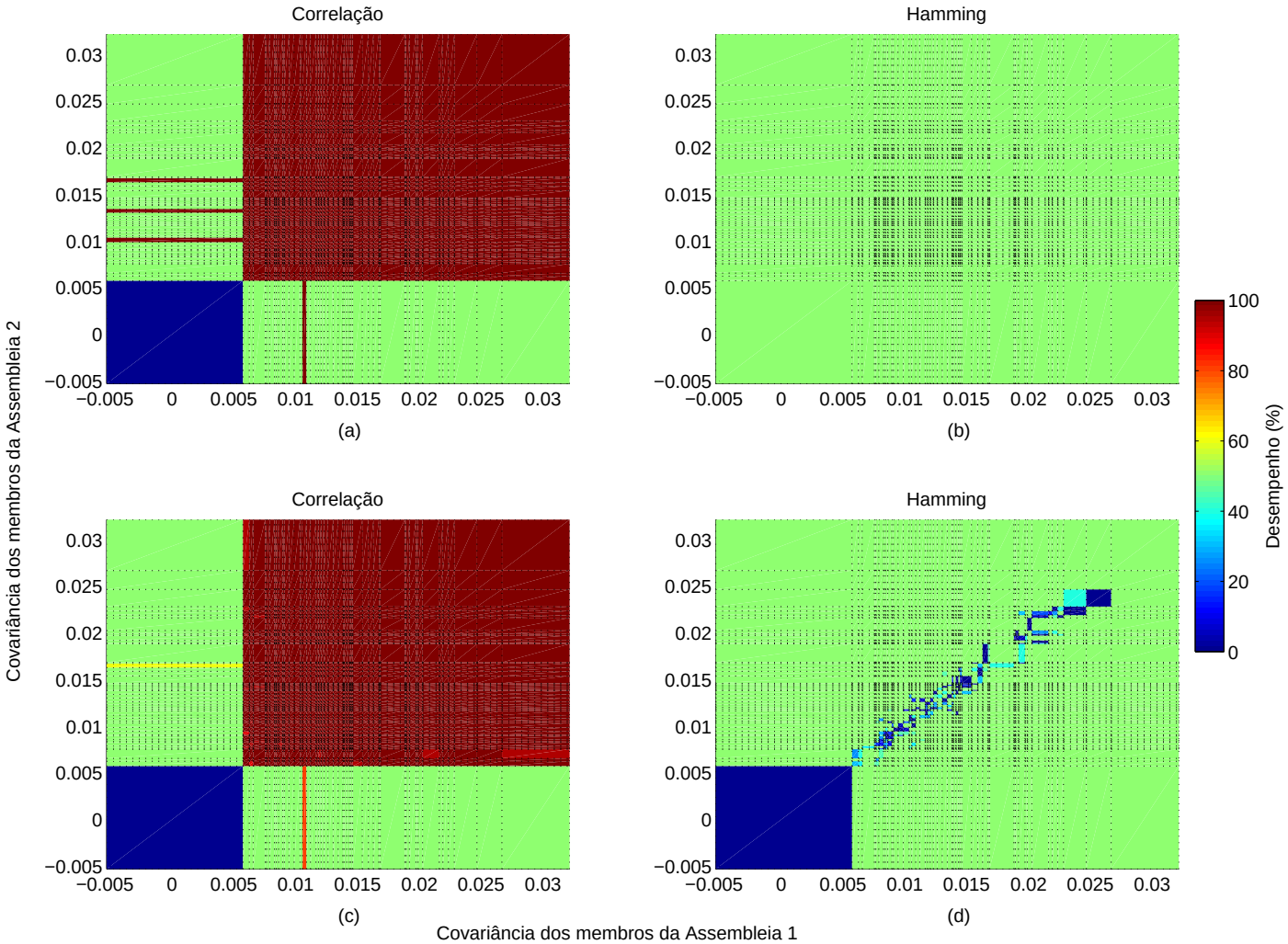
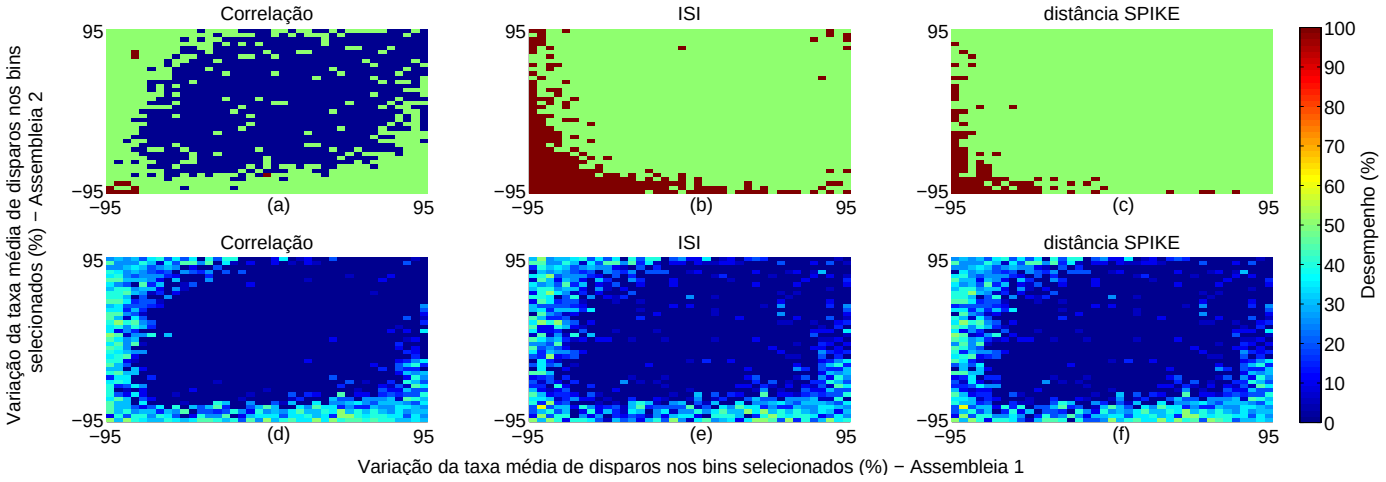


Figura 3.8 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 30 células e ANs com 5 neurônios disjuntos, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator.



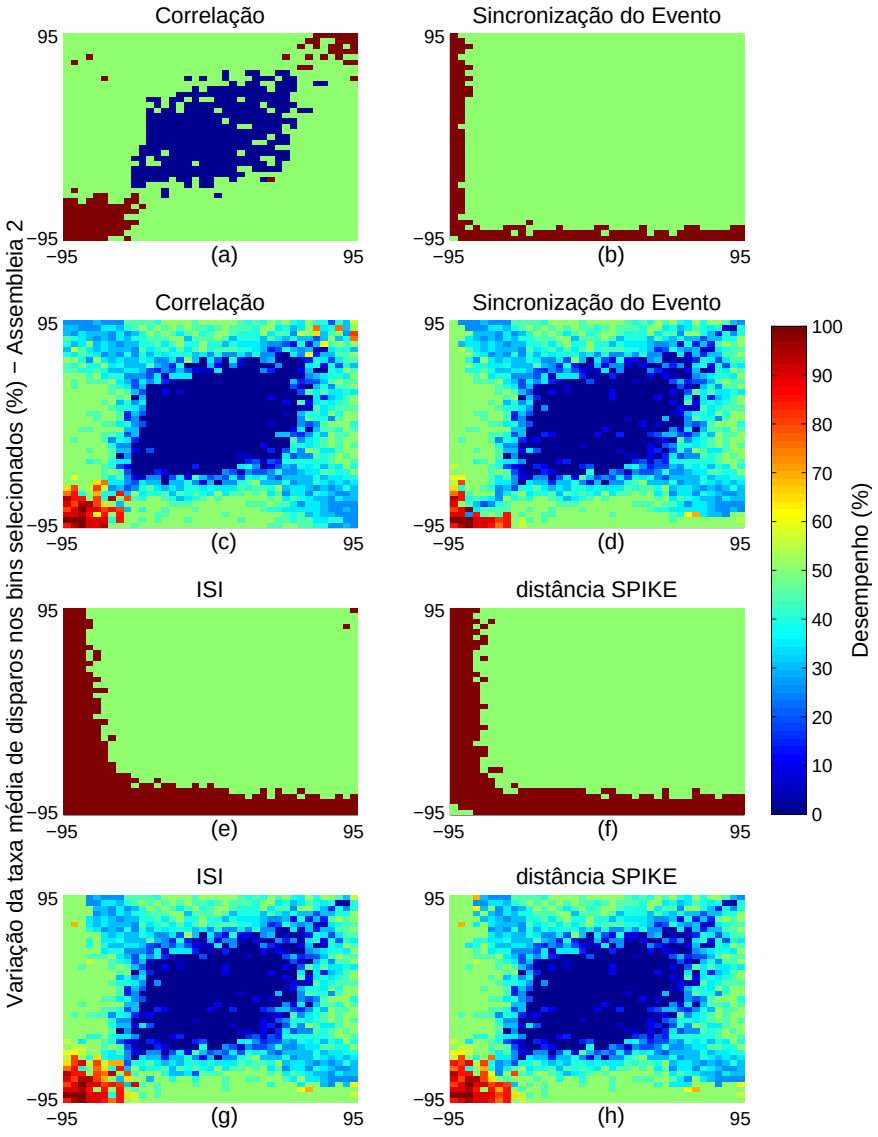
e na terceira linha, exibimos os resultados para a métrica de desempenho relativa ao número de assembleias detectadas (Equação 2.37) e, na segunda e na quarta, à classificação dos neurônios (Equação 2.39). Comparando o resultado ao obtido no cenário com ANs independentes (Figura 3.3), observamos que, para a correlação, ANs deixam de ser detectadas em regiões em que uma AN é de excitação e outra, de inibição (Figura 3.9a), o que acarreta uma queda de desempenho para a classificação dos neurônios nas regiões correspondentes (Figura 3.9c). Já para a Sincronização do Evento, para o ISI e para a distância SPIKE observamos uma ligeira queda na detecção de ANs para algumas regiões de inibição (Figuras 3.9b, 3.9e e 3.9f, respectivamente), o que, consequentemente, é acompanhado por um decréscimo no desempenho para a classificação dos neurônios (Figuras 3.9d, 3.9g e 3.9h, respectivamente).

Os efeitos observados para uma população de 60 células (caso 2, descrito na Seção 3.1.1) foram similares aos obtidos para uma população de 30 células e, portanto, foram omitidos.

3.1.2.2.2 ICA: Simulações utilizando trens de *spike* que consideram a correlação entre neurônios em ANs com ao menos um neurônio exclusivo

As simulações que consideravam a correlação cruzada entre as células em uma janela de 10 ms e numa população de 30 células (caso 1, descrito na Seção 3.1.1) foram exibidas nas

Figura 3.9 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira e na terceira linha, e, na classificação dos neurônios, na segunda e na quarta linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 30 células e ANs com 5 neurônios disjuntos, com 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator.



Figuras 3.10a e 3.10b, referentes à aplicação do ICA às métricas de desempenho relativas ao número de assembleias detectadas (Equação 2.37); e nas Figuras 3.10c e 3.10d, à classificação dos neurônios (Equação 2.39).

Notamos que comparativamente ao cenário em que as ANs eram independentes (Figura 3.7), as duas métricas mantiveram seu desempenho para a detecção do número de ANs (Figuras 3.10a e 3.10b), porém, apresentaram queda na performance para a detecção das células (Figuras 3.10c e 3.10d).

3.1.2.3 ICA: Simulações de ANs sem neurônios disjuntos

Para finalizar o estudo do ICA, aplicamos o algoritmo a ANs sem neurônios disjuntos, descritos nos casos 1 e 2 da Seção 3.1.1. Iniciamos a análise com ANs modeladas em trens de *spike* que utilizam a distribuição de Poisson e, a seguir, a correlação cruzada entre as células.

3.1.2.3.1 ICA: Simulações utilizando trens de *spike* segundo a distribuição de Poisson em ANs sem neurônios disjuntos

De maneira semelhante à Seção anterior, exibimos apenas os resultados nos quais esta topologia apresentou um comportamento diferente em relação ao que já foi apresentado.

Os resultados para os trens artificiais gerados segundo o modelo de Poisson numa população de 30 células (caso 1, descrito na Seção 3.1.1) foram exibidos na Figura 3.11. Na primeira linha exibimos o resultado para a aplicação da métrica de desempenho relativa à detecção de ANs (Equação 2.37) e, na segunda, à classificação dos neurônios (Equação 2.39). Fazendo uma comparação com o cenário de ANs disjuntas, exibido na Figura 3.2, notamos para a correlação (Figuras 3.11a e 3.11d) uma queda no desempenho para a detecção de ANs e para a classificação dos neurônios nas regiões de máxima excitação e inibição da assembleia 1 (laterais esquerda e direita das figuras). Além disso, tanto o ISI quanto a distância SPIKE melhoraram seu desempenho na detecção de ANs para regiões de inibição das duas ANs (Figuras 3.11b e 3.11c), porém, esta melhoria não é acompanhada pela correta classificação dos neurônios (Figuras 3.11e e 3.11f).

A seguir, ainda para a mesma topologia, modificamos artificialmente a taxa média de disparos da população para 64 *spikes/ segundo* para verificar os efeitos desta grandeza na caracterização de ANs. As métricas de similaridade com um desempenho diferente seguem na Figura 3.12, nas linhas ímpares para a detecção de ANs (Equação 2.37) e nas pares para a classificação dos neurônios (Equação 2.39). Foi possível observar, em comparação com o cenário em que a taxa de disparos é obtida experimentalmente (Figura 3.11), uma melhoria apresentada no desempenho da correlação, tanto na detecção de ANs (Figura 3.12a) quanto

Figura 3.10 – Desempenho do ICA associado a diferentes métricas de similaridade quando aplicado a trens artificiais que modelam ANs com 5 neurônios não exclusivos gerados segundo o modelo de correlação entre os neurônios (MACKE *et al.*, 2009) com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Na primeira linha, exibimos o desempenho para o número de assembleias detectadas (Equação 2.37)e, na segunda, para a classificação dos neurônios (Equação 2.4.2). As linhas pontilhadas pretas em cada um dos eixos delimitam os 39 valores distintos de correlação cruzada (extraídos de dados reais) e foram inseridas para salientar que os valores de correlação não são uniformemente distribuídos.

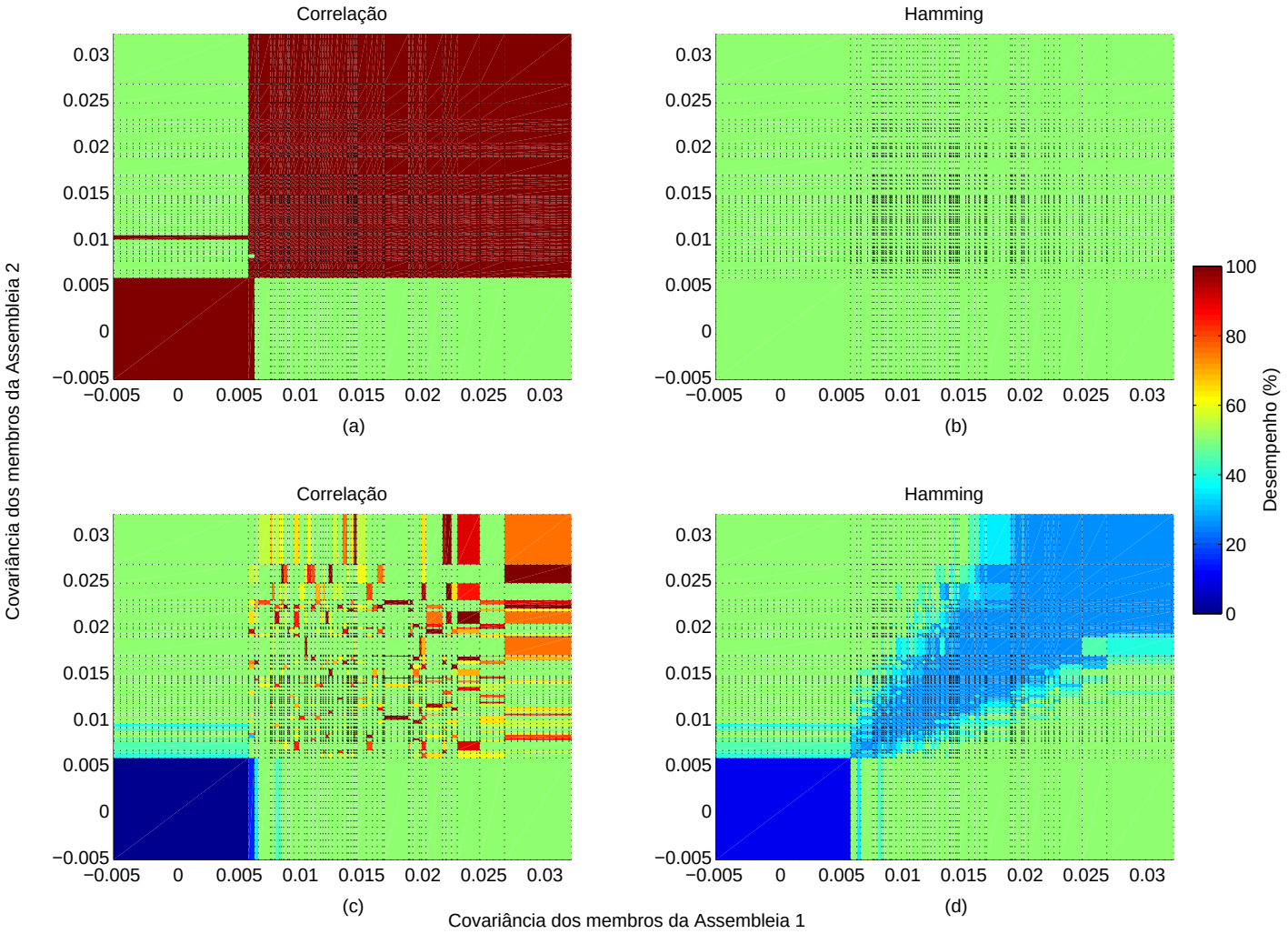
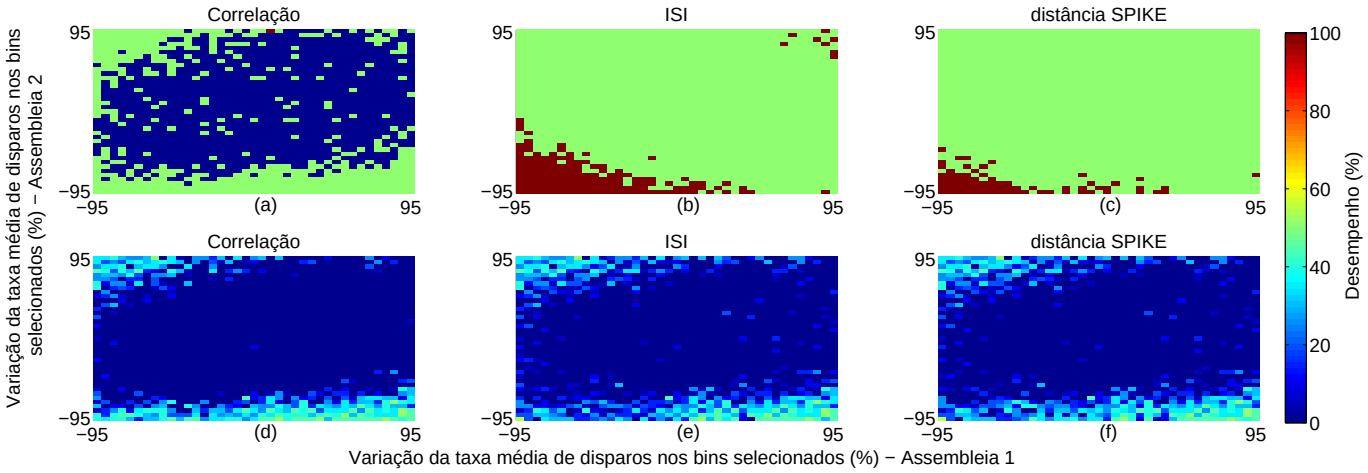


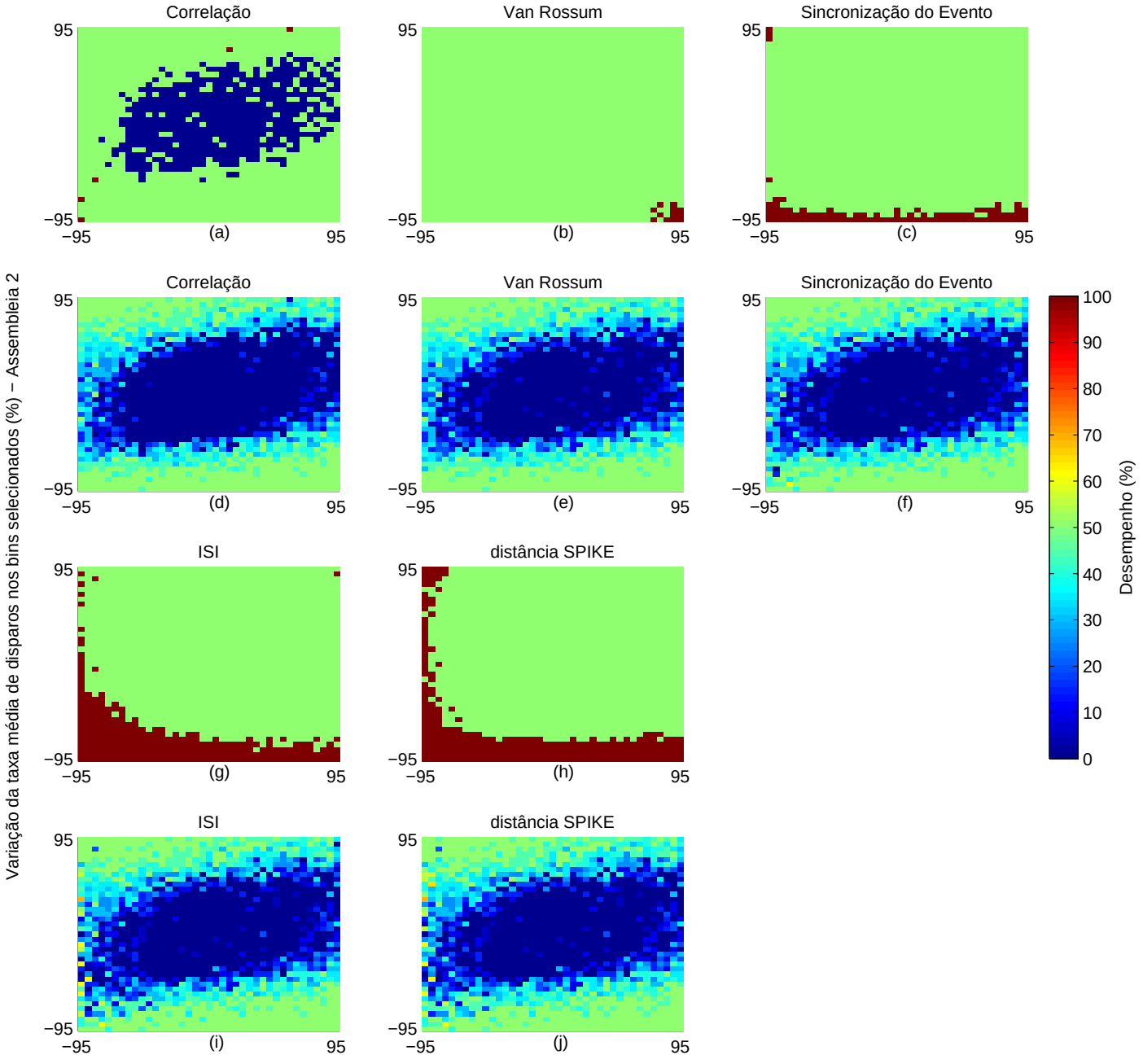
Figura 3.11 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 30 células e ANs sem neurônios disjuntos, sendo a primeira comunidade com 10 neurônios e, a segunda, com 5. Em um total de 10^4 amostras, selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator.



na classificação de neurônios (Figura 3.12d). Além disso, notamos aumento na detecção de ANs para a métrica de Van Rossum em uma pequena região de excitação da assembleia 1 aliada à inibição da assembleia 2 (Figura 3.12b). Já as métricas Sincronização do evento, ISI e distância SPIKE (Figuras 3.12c, 3.12g e 3.12h, respectivamente) mostraram uma melhoria na detecção de comunidades nas regiões de inibição da assembleia 2. No entanto, observamos nas linhas 2 e 4 da Figura 3.12 que a melhoria na detecção das ANs não se refletiu em melhorias no desempenho para a classificação dos neurônios.

A seguir, aumentamos o tamanho da população de células simuladas e retornamos à taxa de disparos experimental, de 22,37 *spikes/segundo*. Novamente, optamos por exibir apenas os resultados com diferenças em relação ao que já foi apresentado. Nas linhas 1 e 3 da Figura 3.13, apresentamos o resultado para a aplicação da métrica de desempenho relativa à detecção de ANs (Equação 2.37) e, nas linhas 2 e 4, à classificação dos neurônios (Equação 2.39). Em comparação ao cenário em que o número de células era menor (Figura 3.11), foi possível observar melhoras na detecção de ANs. Para a correlação (Figura 3.13a), houve uma diminuição na região em que nenhuma assembleia foi detectada. Algumas comunidades passaram a ser observadas para a Sincronização do Evento, quando a AN 1 era de inibição (Figura 3.13b). Tanto o ISI quanto a distância SPIKE melhoraram seu desempenho na detecção de

Figura 3.12 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira e na terceira linha, e, na classificação dos neurônios, na segunda e na quarta linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 30 células e ANs sem neurônios disjuntos, sendo a primeira comunidade com 10 neurônios e, a segunda, com 5. Em um total de 10^4 amostras, selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator.



ANs para regiões de inibição das duas ANs (Figuras 3.13e e 3.13f, respectivamente), mas o ISI apresentou também um bom desempenho na região correspondente à excitação das ANs, no primeiro quadrante da figura. Verificamos também que o aumento no número de células leva a uma melhoria no desempenho para a classificação dos neurônios, mas que o resultado para as diferentes métricas de similaridade é parecido.

Para finalizar, analisamos o cenário em que a população de 60 células, descrita na Seção 3.1.1 tinha sua taxa média de disparos aumentada para 64 *spikes/segundo*. Exibimos na primeira e na terceira linha da Figura 3.14 o resultado referente à aplicação da métrica de desempenho relativa à detecção de ANs (Equação 2.37) e na segunda e na quarta linha à classificação dos neurônios (Equação 2.39). Ao compararmos o resultado obtido com para mesma topologia, porém, com uma menor taxa média de disparos (Figuras 3.13 e 3.14) observamos para a correlação (Figura 3.14a) que ANs começam a ser detectadas nas regiões em que as ANs tem interação de mesma natureza (canto do primeiro e do terceiro quadrante). Além disso, para as métricas de Victor e Purpura (Figura 3.14b) e de Van Rossum (Figura 3.14c), assembleias foram detectadas nos pontos extremos em que uma das ANs possuía interações de inibição e a outra, de excitação. Houve uma melhoria na detecção de ANs quando a AN1 era inibitória para as métricas de Sincronização do Evento, ISI e distância SPIKE (Figuras 3.14g, 3.14h e 3.14i, respectivamente). Verificamos também uma melhoria no desempenho para a classificação dos neurônios (segunda e quarta linhas da Figura 3.14), sem diferenças significativas entre as métricas de similaridade.

Observamos também que houve uma ligeira mudança no comportamento do ISI em relação ao cenário em que as ANs eram independentes (Figura 3.5n). Anteriormente, o melhor desempenho obtido foi quando uma das ANs possuía uma interação de inibição, e a outra, de excitação. Agora, foi observado quando as duas ANs possuíam interações de mesma natureza.

3.1.2.3.2 ICA: Simulações utilizando trens de *spike* que consideram a correlação entre neurônios em ANs sem neurônios disjuntos

Para finalizar a análise do ICA, simulamos uma população de 30 células sem neurônios disjuntos (caso 1, apresentado na Seção 3.1.1) que considerava a correlação cruzada entre as células em janelas de 10 ms. O desempenho da aplicação do ICA referente à métrica de detecção de ANs (Equação 2.37) é apresentado na primeira linha da Figura 3.15 enquanto que, para a classificação dos neurônios (Equação 2.39), na segunda linha.

Observamos que, comparativamente ao cenário em que as ANs eram independentes (Figura 3.7), houve uma queda no desempenho para a identificação das ANs. No entanto, ainda existiram regiões em que as duas ANs foram detectadas, em contraste ao resultado

Figura 3.13 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira e na terceira linha, e na classificação dos neurônios (Seção 2.4.2), na segunda e na quarta, quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 60 células e ANs sem neurônios disjuntos, sendo a primeira comunidade com 25 neurônios e, a segunda, com 10. Em um total de 10^4 amostras, selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator.

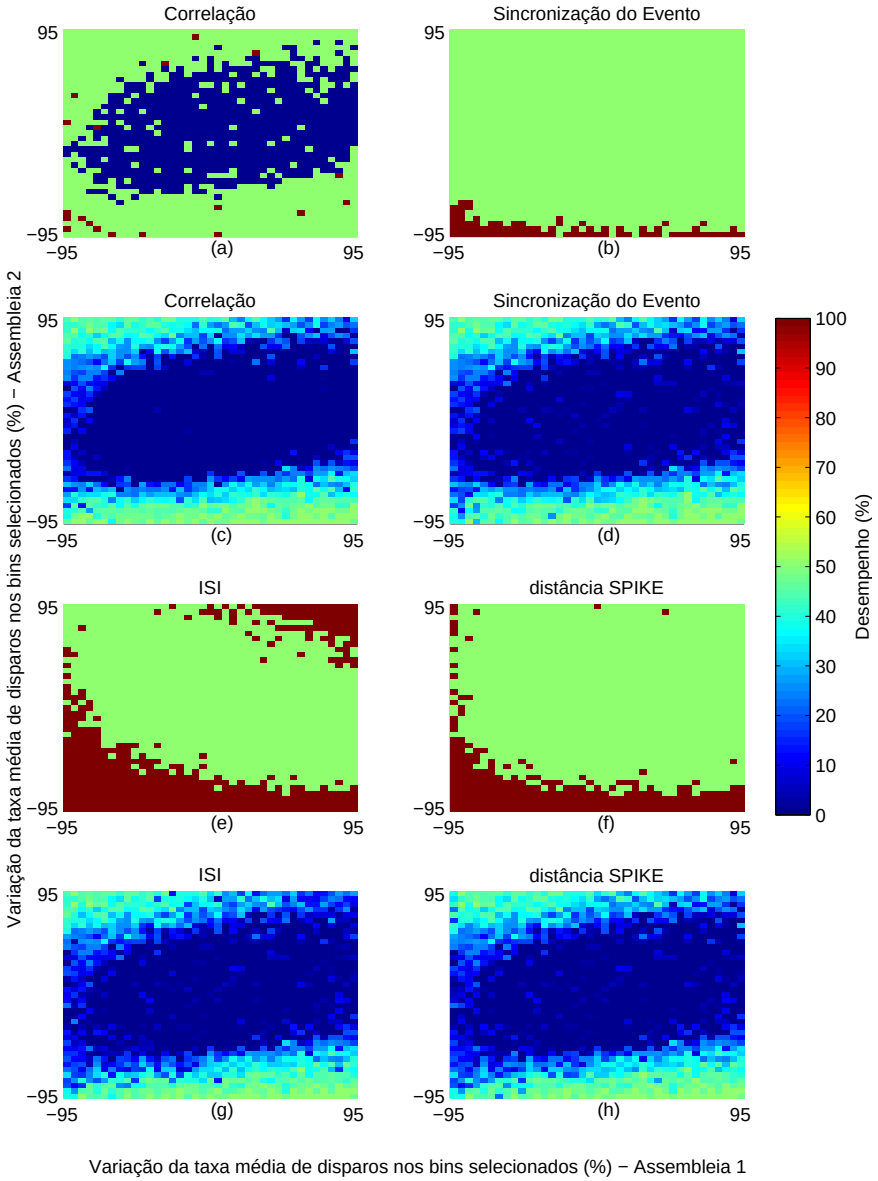


Figura 3.14 – Desempenho do ICA associado a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira e na terceira linha, e na classificação dos neurônios (Seção 2.4.2), na segunda e na quarta, quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células e ANs sem neurônios disjuntos, sendo a primeira comunidade com 25 neurônios e, a segunda, com 10. Em um total de 10^4 amostras, selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator.

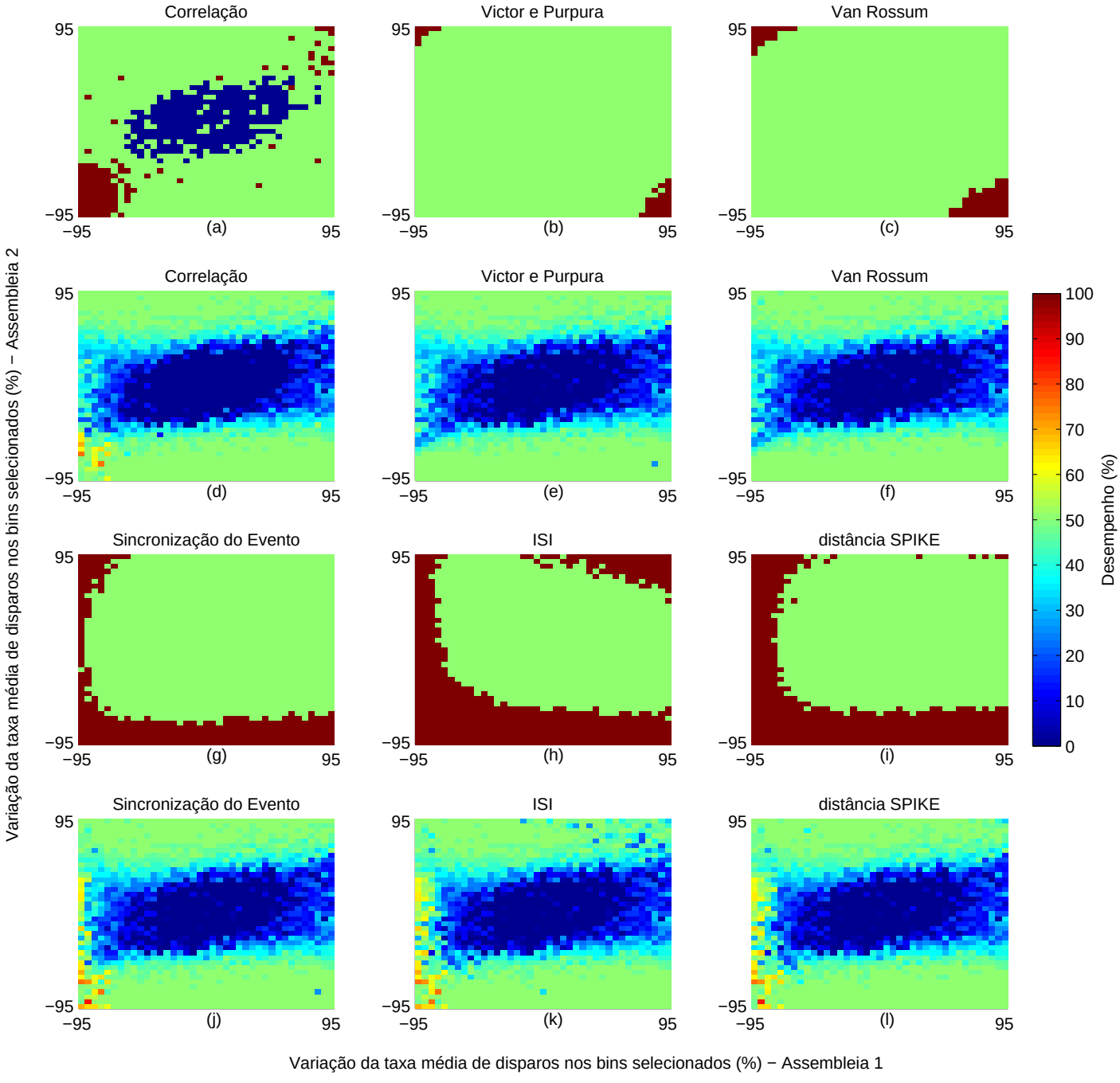
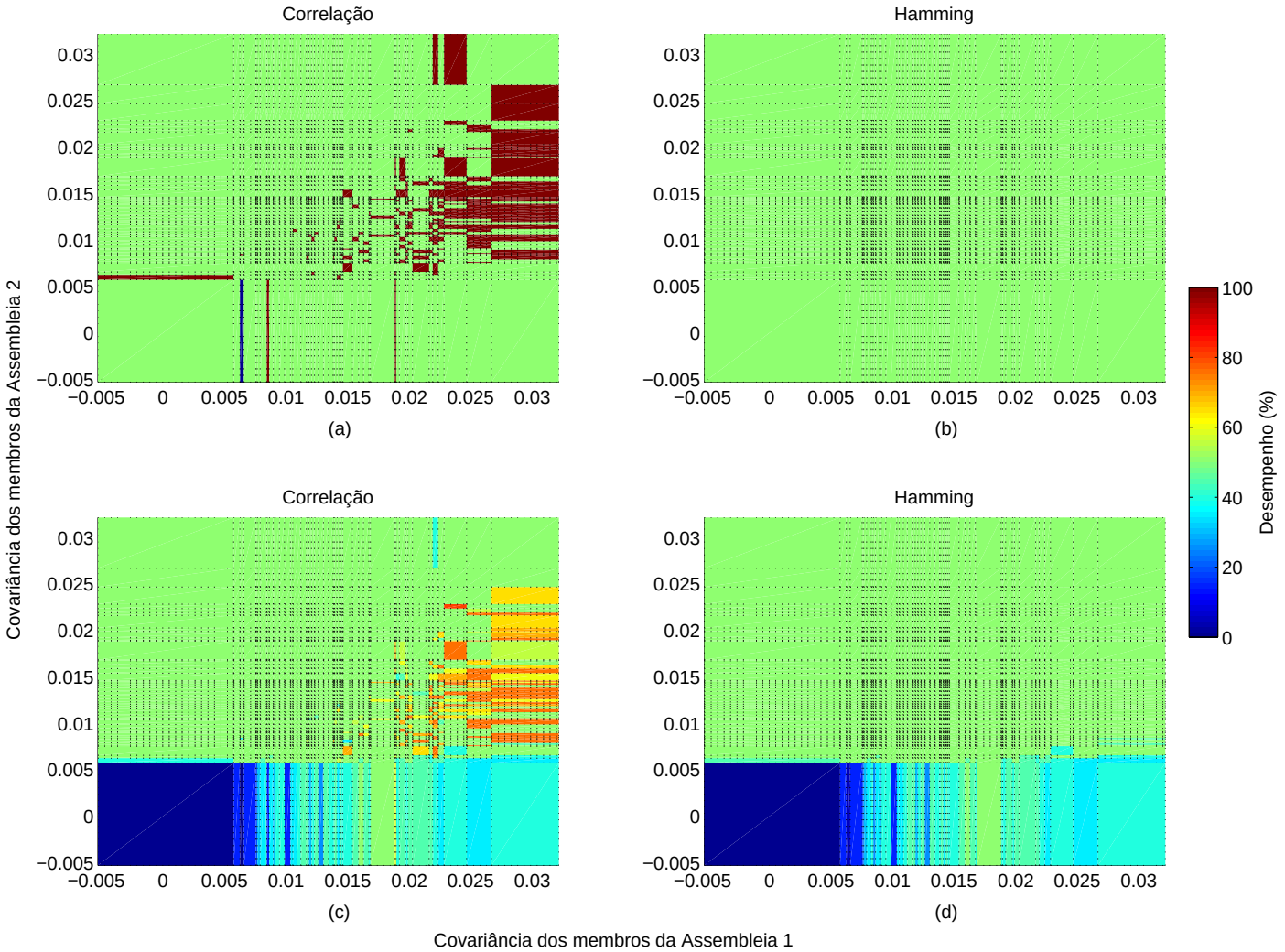


Figura 3.15 – Desempenho do ICA associado a diferentes métricas de similaridade quando aplicado a trens artificiais que modelam ANs sem neurônios disjuntos gerados segundo o modelo de correlação entre os neurônios (MACKE *et al.*, 2009) com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Na primeira linha, exibimos o desempenho para o número de assembleias detectadas (Equação 2.37) e, na segunda, para a classificação dos neurônios (Equação 2.4.2). As linhas pontilhadas pretas em cada um dos eixos delimitam os 39 valores distintos de correlação cruzada (extraídos de dados reais) e foram inseridas para salientar que os valores de correlação não são uniformemente distribuídos.



obtido para trens simulados utilizando a distribuição de Poisson (Figura 3.11a). Para a classificação dos neurônios, houve uma queda no desempenho para as duas métricas de similaridade (Hamming e correlação) no ponto que simulava interações inibitórias entre as células da assembleia 2. No entanto, como apenas um valor de correlação negativa foi medido na escala temporal de 10 ms não consideramos essa queda de desempenho para efetuar conclusões.

3.1.3 Resultados: Modularidade

Nesta seção analisamos o desempenho do algoritmo Modularidade aplicado aos mesmos conjuntos de dados descritos ao longo da Seção 3.1.2, também utilizando diferentes métricas de similaridade. Simulamos a atividade de ANs independentes, com ao menos um neurônio exclusivo e sem neurônios disjuntos utilizando tanto dados gerados segundo o modelo de Poisson quanto a partir do modelo de Macke *et al.* (2009), que considerava a correlação cruzada entre os membros de uma AN.

3.1.3.1 Modularidade: Simulações de ANs independentes

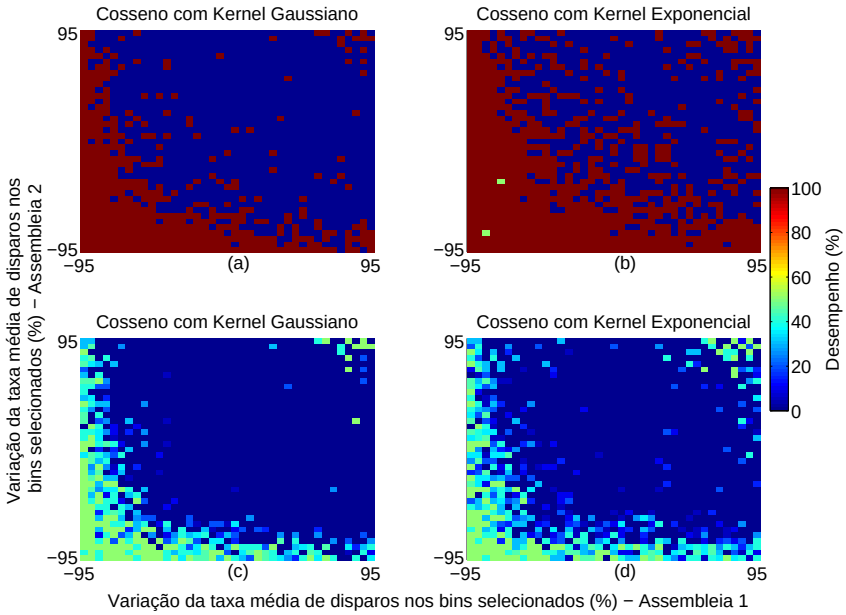
Para a análise da resposta do algoritmo aplicado a ANs independentes, iniciamos com os trens simulados segundo a distribuição de Poisson e, a seguir, passamos para as simulações que contemplavam também a covariância cruzada entre os neurônios.

3.1.3.1.1 Modularidade: Simulações utilizando trens de *spike* segundo a distribuição de Poisson em ANs independentes

Inicialmente, analisamos o resultado para a população simulada de 30 células com ANs independentes (caso 1, descrito na Seção 3.1.1), na taxa média de disparos obtida experimentalmente, de 22,37 *spikes/segundo*. Observamos que, dentre todas as métricas de similaridade avaliadas neste trabalho, apenas o Cosseno retornou um desempenho não nulo para esta topologia. Os resultados foram exibidos na Figura 3.16, a seguir. Na primeira linha apresentamos o resultado para a detecção de ANs (Equação 2.37) e na segunda, para a classificação dos neurônios (Equação 2.39).

Notamos que o uso do *kernel* Gaussiano na detecção de ANs (Figura 3.16a) gerou um menor número de falsos positivos nas regiões centrais, de acordo com o esperado, já que nestes cenários as interações entre as ANs são fracas e presumíamos que nenhuma comunidade seria detectada. Já o *kernel* Exponencial exibiu mais sensibilidade na detecção de ANs nas regiões de inibição (Figura 3.16b). No entanto, o desempenho dos diferentes *kernels* na classificação dos neurônios foi similar (Figuras 3.16c e 3.16d). Além disso, notamos que apesar de um particionamento no número correto de comunidades, houve saturação do desempenho que

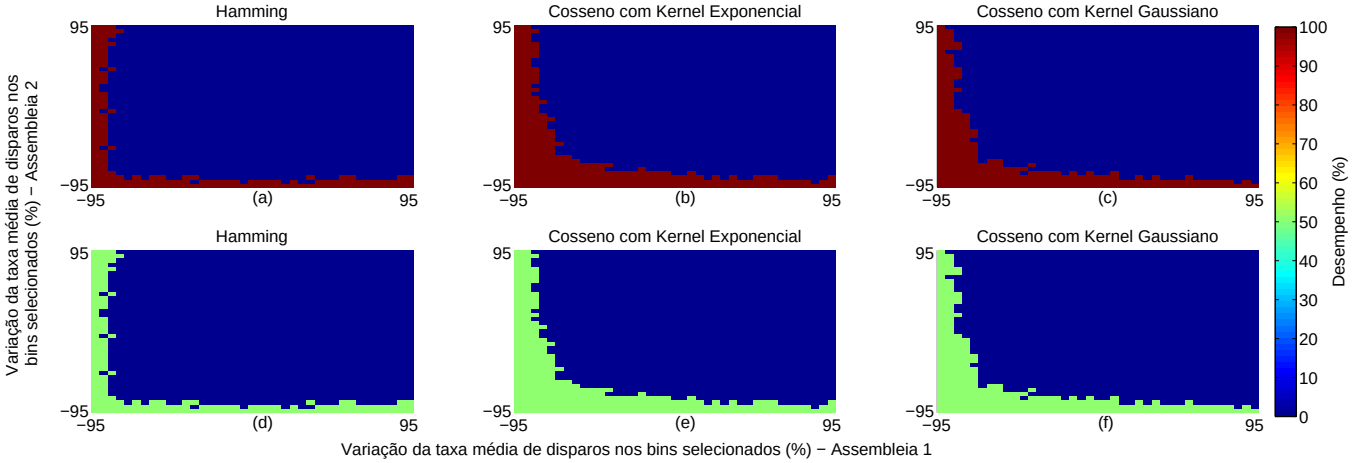
Figura 3.16 – Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios (Seção 2.4.2), na segunda linha, quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 30 células e ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.



não ultrapassou o valor nominal de 50%. Notamos também que, diferentemente do ICA, para a Modularidade não ocorreram detecções espúrias de ao menos uma AN quando os neurônios eram independentes entre si.

A seguir, analisamos o conjunto de dados em que a taxa média de disparos foi elevada para 64 *spikes/segundo*, ainda em uma população de 30 células. O desempenho foi apresentado na Figura 3.17, a seguir. Exibimos apenas o resultado para as métricas com desempenho não nulo. Na primeira linha, observa-se o desempenho para a detecção de ANs (Equação 2.37) e, na segunda, para a classificação dos neurônios (Equação 2.39). Comparativamente ao conjunto de dados simulados com uma taxa média de disparos mais baixa (Figura 3.16), notamos que a associação com a métrica de Hamming passou a detectar ANs em regiões de interação inibitória para as assembleias 1 e 2 (Figura 3.17a). Além disso, o desempenho para o uso do Cosseno na identificação de ANs se modifica, tornando-se mais nítido, sem detecções espúrias nas situações em que as células são independentes entre si (Figuras 3.17b e 3.17c). O desempenho para a classificação de neurônios continuou saturado em 50% (Figuras 3.17d, 3.17e e 3.17f).

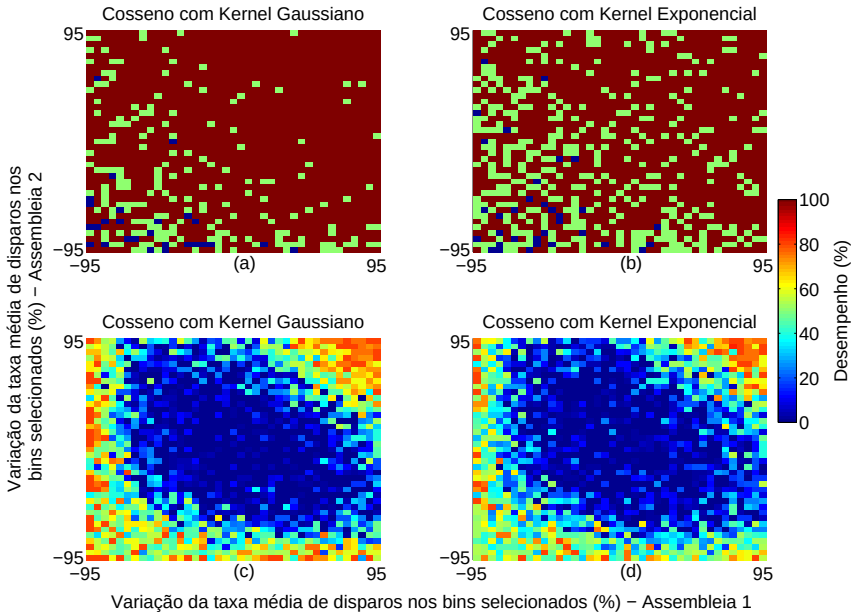
Figura 3.17 – Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 30 células e ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator.



Na próxima etapa verificamos o efeito de elevarmos o número de células e, consequentemente, o tamanho das ANs (caso 2 descrito na Seção 3.1.1), porém, novamente na taxa média de disparos obtida experimentalmente, de 22,37 *spikes/segundo*. Exibimos na Figura 3.18 o resultado da combinação da Modularidade com o Cosseno, pois foi a única métrica de similaridade que apresentou desempenho não nulo. Foi possível observar a detecção de ANs mesmo num cenário em que as células eram independentes (parte central das Figuras 3.18a e 3.18b). Para a classificação dos neurônios, nas Figuras 3.18c e 3.18d, observamos um baixo desempenho nas regiões centrais e que o desempenho máximo foi de 80%.

Analisamos a seguir, novamente o efeito do aumento da taxa média de disparos a 64 *spikes/segundo*, porém em uma população maior com ANs maiores (caso 2 descrito na Seção 3.1.1). Na Figura 3.19, a seguir, a primeira linha exibiu o desempenho das métricas de similaridade associadas à Modularidade na detecção de ANs (Equação 2.37) e a segunda, à classificação dos neurônios (Equação 2.39). Ao compararmos com o cenário que as ANs eram menores (Figura 3.17), observamos um aumento no desempenho tanto para a detecção de ANs quanto para a classificação dos neurônios. Notamos que o desempenho para a classificação dos neurônios, quando não nulo, foi numericamente igual a 80% em praticamente todos os cenários simulados. Além disso, foi possível notar que a métrica de Hamming (Figuras 3.19a e 3.19d) exibiu um desempenho superior em relação ao Cosseno (Figuras 3.19b, 3.19c,

Figura 3.18 – Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator.



3.19e e 3.19f) na região de excitação de uma assembleia e inibição da outra (segundo e quarto quadrantes das figuras). Por sua vez o desempenho foi inferior para as interações de excitação das duas ANs (primeiro quadrante das figuras).

Para finalizar, analisamos o efeito que atrasos de fase provocaram no desempenho das métricas de similaridade associadas à Modularidade. Optamos por continuar utilizando uma população de 60 células, com ANs maiores (caso 2 apresentado na Seção 3.1.1) com taxa média de disparos de 64 *spikes/segundo*, pois, neste caso, as diferenças foram mais acentuadas. Cinco neurônios de uma das ANs tiveram sua atividade atrasada em um *bin*. Na primeira linha da Figura 3.20, exibimos o desempenho para a detecção de ANs (Equação 2.37) e, na segunda, para a classificação dos neurônios (Equação 2.39). Em comparação com o caso sem atraso, exibido na Figura 3.19, observamos uma queda de desempenho para a classificação dos neurônios na presença de atrasos. Para a métrica de Hamming a queda de desempenho ocorreu regiões em que as interações eram inibitórias (Figura 3.20d), enquanto que para o Cosseno (Figuras 3.20e e 3.20f), nas regiões de interação excitatória.

Figura 3.19 – Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa média de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.

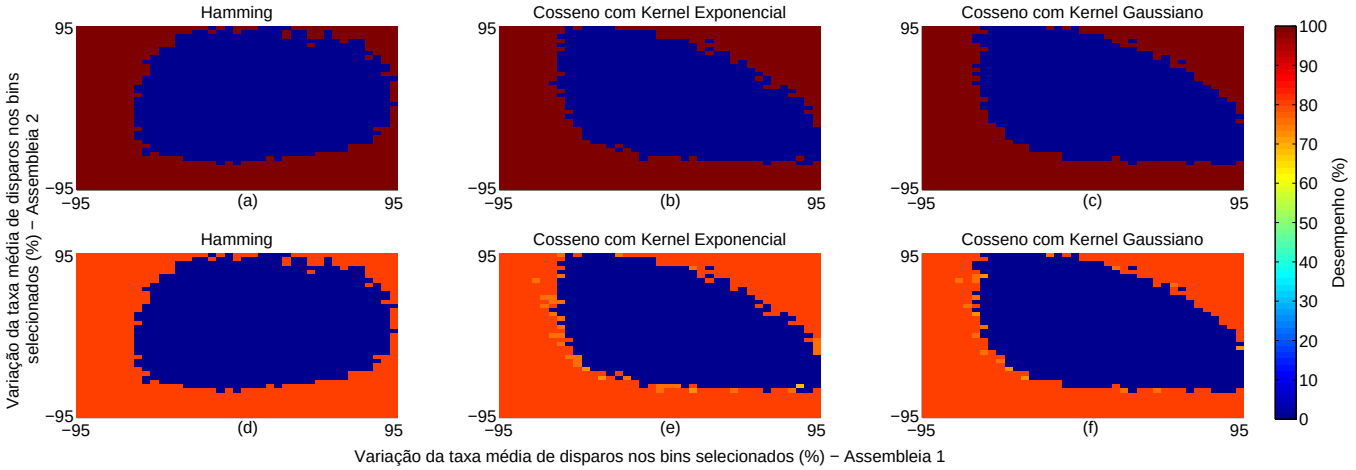


Figura 3.20 – Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células e ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator. A atividade de 5 neurônios de uma das ANs foi atrasada um *bin*.

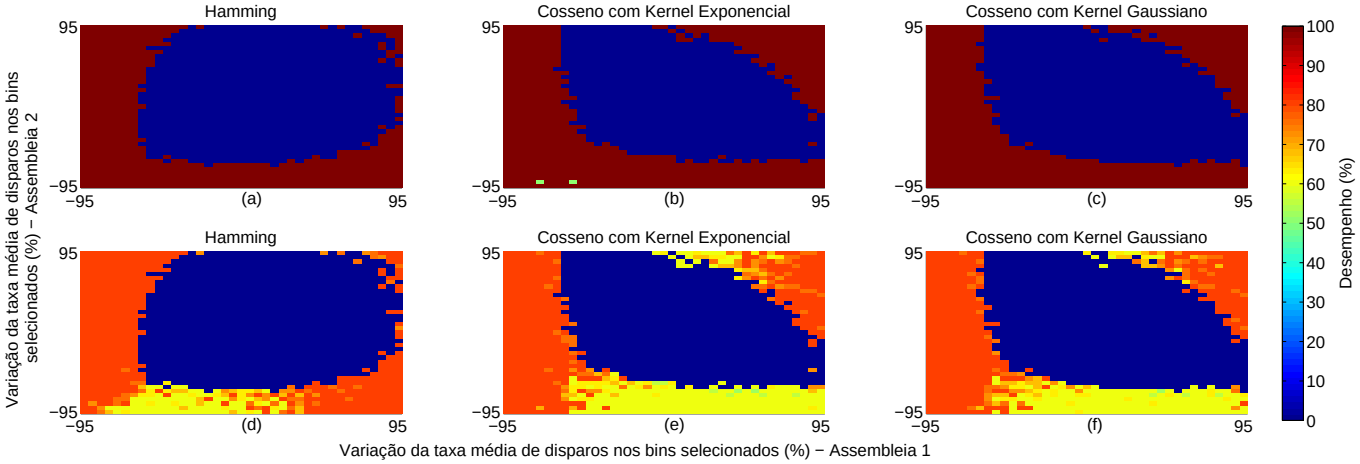
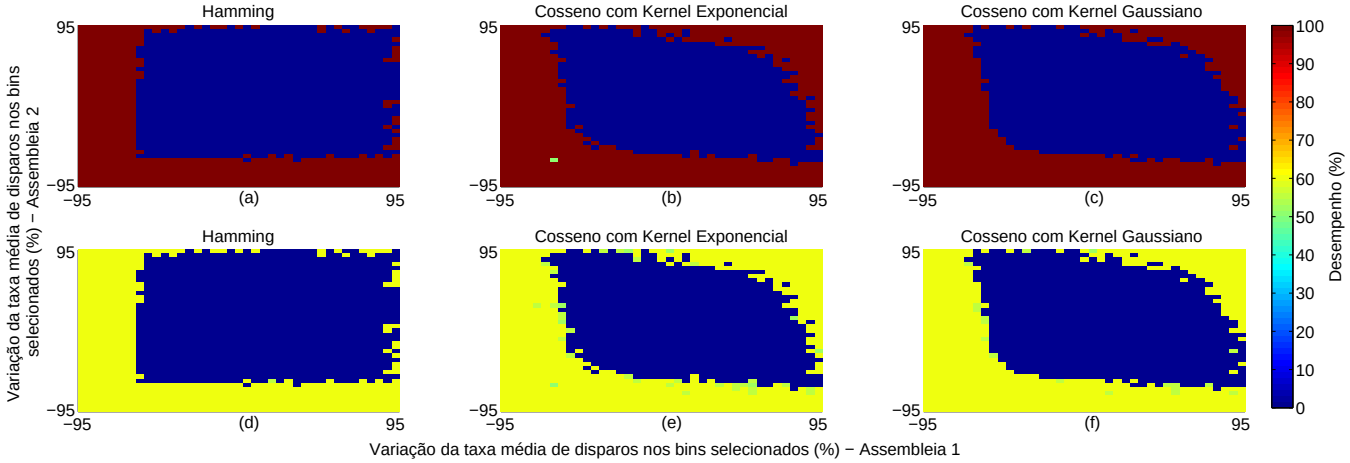


Figura 3.21 – Desempenho da Modularidade associada a diferentes métricas de similaridade na detecção do número de ANs (Seção 2.4.1), na primeira linha, e, na classificação dos neurônios, na segunda linha (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células com ANs com 5 neurônios em comum. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa média de disparos do modelo de Poisson foi escalonada por um fator.



A simulação da população de 30 células, com duas ANs que consideravam as correlações cruzadas entre as células em uma janela de 10 *ms* (caso 1, descrito na Seção 3.1.1) apresentaram desempenho nulo ao associarmos a Modularidade com as métricas de similaridade correlação e Hamming.

3.1.3.2 Modularidade: Simulações de ANs que compartilham neurônios

A Modularidade possui limitações ao trabalhar com elementos que pertencem a mais de uma comunidade, como será discutido na Seção 4.1.2. Ilustraremos esta limitação pela simulação de ANs que compartilham neurônios apenas para a população de 60 células e taxa média de disparos de 64 *spikes/segundo* utilizando o modelo de Poisson (caso 2, descrito na Seção 3.1.1), já que neste cenário as variações de desempenho se tornam mais nítidas.

A Figura 3.21 exibe na primeira linha o desempenho de acordo com a métrica de detecção de ANs (Equação 2.37) e, na segunda, de acordo com a classificação dos neurônios (Equação 2.39). Ao compararmos este cenário com o obtido para ANs independentes, exibido na Figura 3.19, notamos para a métrica de Hamming uma ligeira alteração no desempenho de detecção de ANs, mas, além disso, observamos que o desempenho deixa de atingir o patamar de 80% e satura em 60%. Isso ilustra a limitação do algoritmo, em sua proposta original, de

lidar com populações com elementos em comum.

3.1.4 Resultados: Agrupamento Hierárquico

Apresentamos nesta seção os resultados da aplicação do algoritmo de Agrupamento Hierárquico às métricas de similaridade descritas na Seção 2.2 para trens gerados segundo o modelo de Poisson e para o modelo utilizando a correlação cruzada entre os neurônios. Além disso, conforme descrevemos na Seção 2.3.4, avaliamos diferentes variações do algoritmo, com maneiras distintas de realizar o *Linkage* entre os objetos de um dendrograma (ligações simples, completa, média simples e média ponderada) além de dois índices de validação que visavam o particionamento automático em diferentes comunidades (índices de Davies-Bouldin e Silhouette).

3.1.4.0.1 Agrupamento Hierárquico: Simulações utilizando trens de *spike* segundo a distribuição de Poisson em ANs independentes

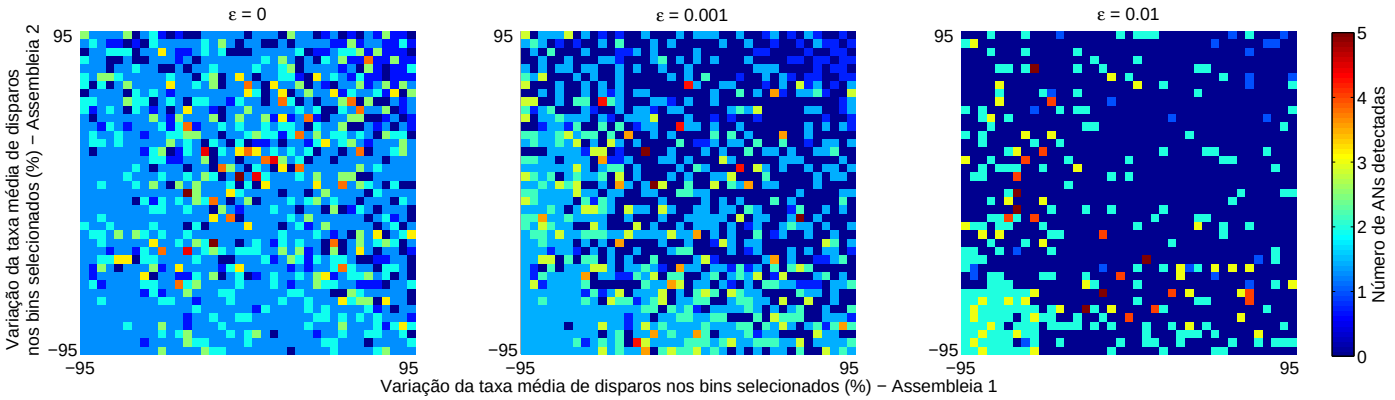
Conforme descrito nas Seções 2.3.4.4.1 e 2.3.4.4.2, as equações utilizadas para definir uma mudança na curvatura recebiam o valor ϵ como parâmetro. Inicialmente, utilizamos o valor de $\epsilon = 0$ tanto para o índice de Davies-Bouldin (Seção 2.3.4.4.1) quanto para o Silhouette (Seção 2.3.4.4.2). Observamos a detecção de um grande número de ANs espúrias em regiões em que as células eram independentes entre si.

Em alguns casos, como o exemplificado na Figura 3.22, a seguir, notamos que o desempenho não se mostrou robusto a pequenas perturbações neste parâmetro. Aqui exibimos o número de ANs detectadas ao utilizarmos o conjunto de dados simulados para a população de 30 células, com taxa média de disparos de 64 *spikes/segundo*, ANs independentes, utilizando a modelagem de Poisson. Além disso, empregamos a correlação como métrica de similaridade, a ligação simples e o índice de validação de Silhouette.

Existiram algumas métricas de similaridade que não apresentaram tamanha variabilidade com pequenas variações de ϵ, porém, pudemos observar que o desempenho estava mais associado à separatividade das comunidades com base no dendrograma do que aos índices de validação. Assim, exibimos a seguir os resultados de uma análise visual das propriedades dos dendrogramas resultantes de diferentes métricas de similaridade e formas distintas de *Linkage*.

Nas próximas figuras exibimos cinco situações distintas, dispostas espacialmente de maneira a representar sua posição no diagrama da Figura 3.1. À esquerda e no canto superior exibimos um dendrograma resultante de interações de inibição para a assembleia 1 e de excitação para a assembleia 2. Por sua vez, no canto superior direito expomos um cenário em

Figura 3.22 – Representação esquemática do efeito da resolução na determinação de comunidades.



que as duas ANs tinham interações excitatórias. Na região central todos os neurônios eram independentes e assim por diante.

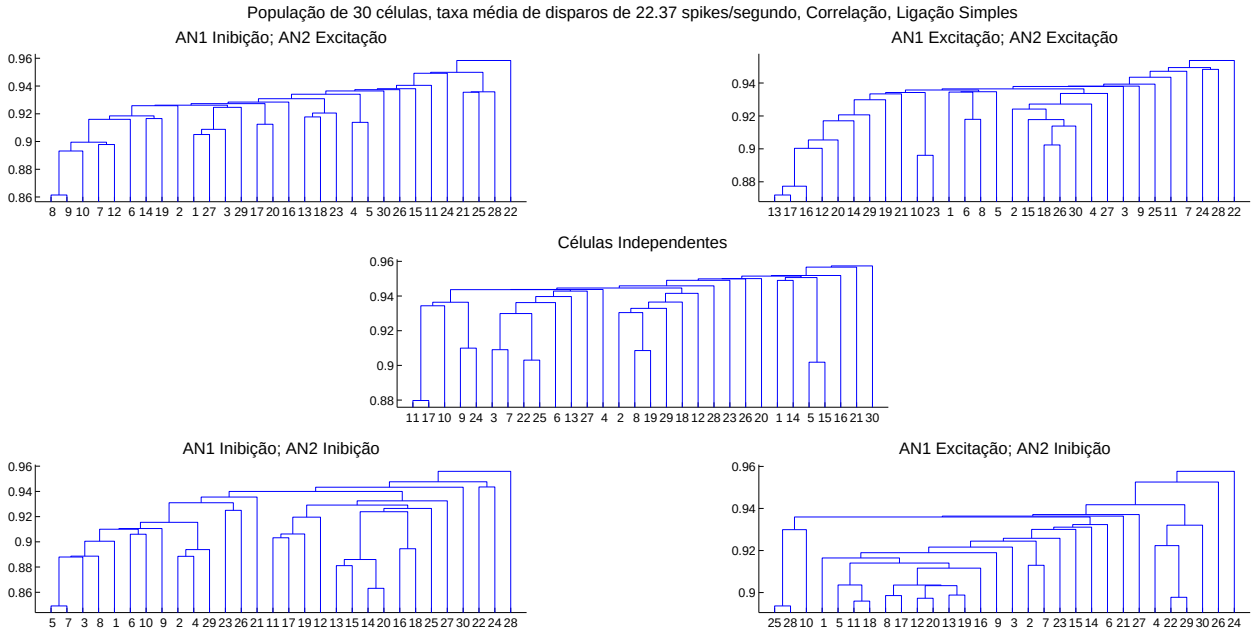
O primeiro parâmetro que analisamos foi a variação dos dendrogramas dependendo da taxa média de disparos da população. Exibimos na Figura 3.23, a seguir, diferentes dendrogramas obtidos ao utilizarmos a Correlação como métrica de similaridade (já aplicando a adaptação descrita na Seção 2.3.6) e a ligação do tipo simples com uma taxa média de disparos de 22,37 *spikes/segundo*. Na Figura 3.24, exibimos o resultado do mesmo cenário para 64 *spikes/segundo*.

Foi possível ressaltar a dificuldade para identificarmos visualmente a existência de comunidades no caso em que a taxa de disparos é baixa (Figura 3.23). Notamos que os neurônios correspondentes às comunidades pré-definidas tem uma tendência a se reunirem, porém, a diferença em relação ao cenário de células independentes não foi clara e ainda existiam associações espúrias com células independentes. Ressaltamos que os cenários exibidos nas figuras correspondem aos valores limites de excitação e inibição (cantos da Figura 3.1) e, portanto, a diferenciação em cenários intermediários foi ainda mais difícil.

Já para um aumento na taxa de disparos (Figura 3.24), observamos que a delimitação das comunidades fica mais evidente, sendo possível diferenciá-las de células independentes.

Na próxima etapa, analisamos a resposta às diferentes formas de *Linkage*. Exibimos os resultados para a taxa média de disparos de 64 *spikes/segundo*, cenário em que as diferenças foram mais evidentes. Continuamos exibindo os resultados para o uso da Correlação como métrica de similaridade. Comparamos um dos cenários em que a população simulada tinha uma atividade independente (Figura 3.25) com um cenário em que uma das ANs tinha interações de excitação e, a outra, de inibição (Figura 3.26).

Figura 3.23 – Dendrogramas resultantes do uso da correlação com a ligação do tipo simples para uma população de 30 células com taxa média de disparos de 22,37 *spikes/segundo* e ANs independentes. Os tipos de interação entre as ANs variam conforme descrito na legenda.

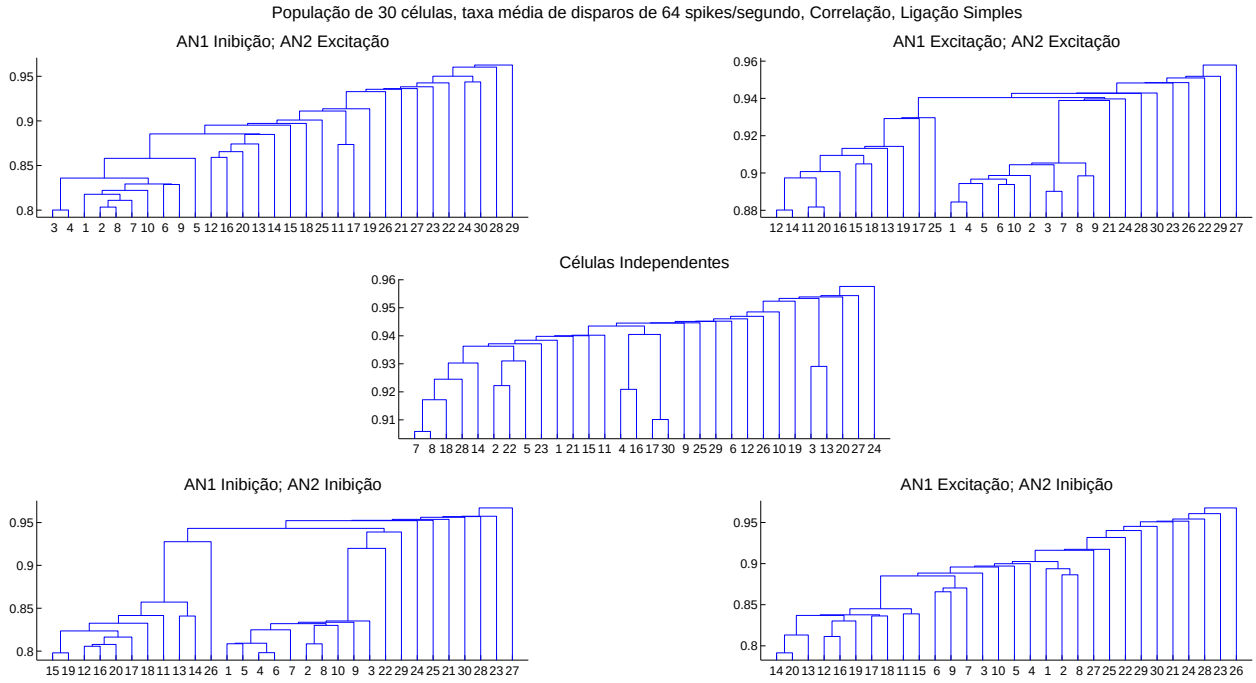


Observamos que tanto a ligação do tipo simples quanto a média simples foram as que apresentaram diferenças mais evidentes entre o cenário em que as ANs estão presentes versus uma população independente. Nestes casos, verificamos no dendrograma um aumento na diferença entre a folha que reúne os elementos da comunidade e a próxima folha. Para a ligação simples a diferença entre folhas foi menos expressiva que para a média simples, no entanto, a primeira possibilitou a diferenciação entre células independentes, representada pelos elementos conectados em cascata, e as comunidades, que foram agrupados nas folhas.

Para a ligação completa e para a média ponderada a diferença entre cenários com ANs e células independentes foi menos perceptível visualmente, apesar de as distâncias entre os elementos terem limiares distintos no eixo vertical das figuras. A identificação do ponto ideal de particionamento do dendrograma mostrou-se difícil, e só foi possível devido ao conhecimento prévio da estrutura dos dados. A ligação média ponderada apresentou uma estrutura mais apropriada que a ligação completa na identificação de células independentes, já que a diferença em relação às ANs foi um pouco mais evidente. Notamos também que, para todas as formas de *Linkage*, foi mais fácil visualizar as assembleias com interações inibitórias.

A seguir, comparamos o comportamento das demais métricas de similaridade anali-

Figura 3.24 – Dendrogramas resultantes do uso da correlação com a ligação do tipo simples para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Os tipos de interação entre as ANs variam conforme descrito na legenda.



sadas neste trabalho quando utilizadas na confecção de dendrogramas. Exibimos apenas os dendrogramas das ligações simples (Figuras 3.27 e 3.28) e média simples (Figuras 3.29 e 3.30), que se mostraram mais promissoras. Continuamos com a estratégia de exibir inicialmente o resultado para um dos cenários em que a população era independente (Figuras 3.27 e 3.29), seguida por um cenário em que uma das ANs tinha interações de excitação e a outra, de inibição (Figuras 3.28 e 3.30).

Foi possível notar que, nesta taxa média de disparo, a métricas de Hamming não agrupava corretamente as ANs nem ao utilizarmos a ligação simples nem a média simples. Para o uso do cosseno com os *kernels* Gaussiano e Exponencial, notamos para a ligação simples que a AN excitatória começa a se destacar, já que seus nós não formam folhas, porém houve dificuldade na diferenciação visual das ANs sem um conhecimento prévio da estrutura dos dados. Não observamos uma estruturação do dendrograma do cosseno ao utilizarmos a ligação média simples. Para a Sincronização do evento, tanto para a ligação simples quanto para a média simples observamos o início do agrupamento das células que compõem as ANs, porém ainda existem neurônios independentes incorretamente associados às assembleias e a

Figura 3.25 – Dendrogramas resultantes do uso da correlação com diferentes tipos de *Linkage* para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Células Independentes.

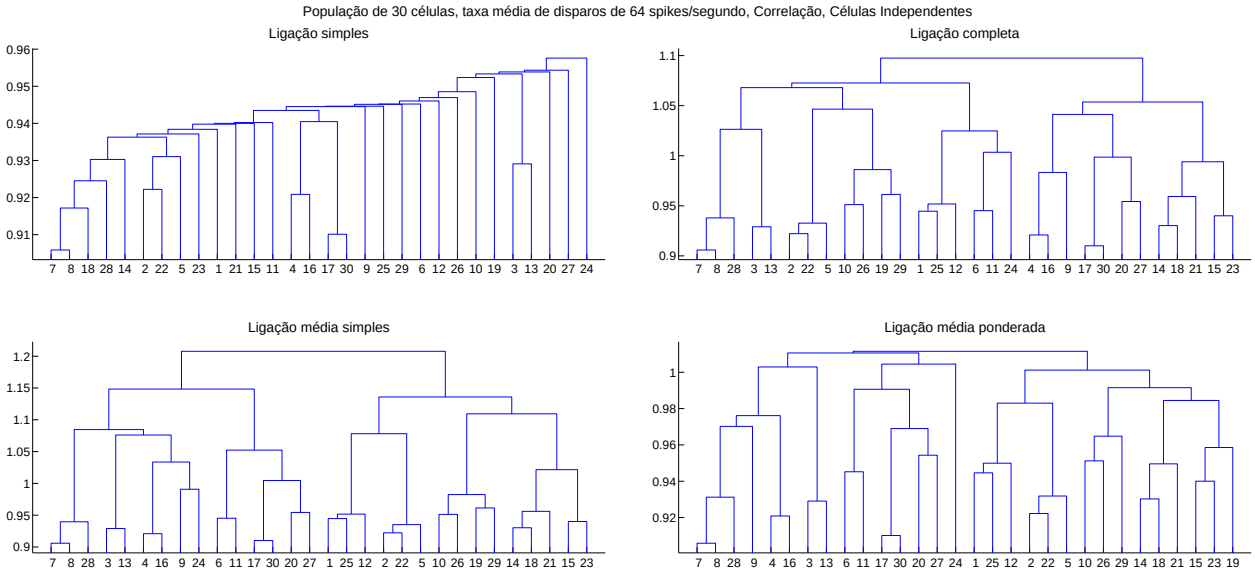
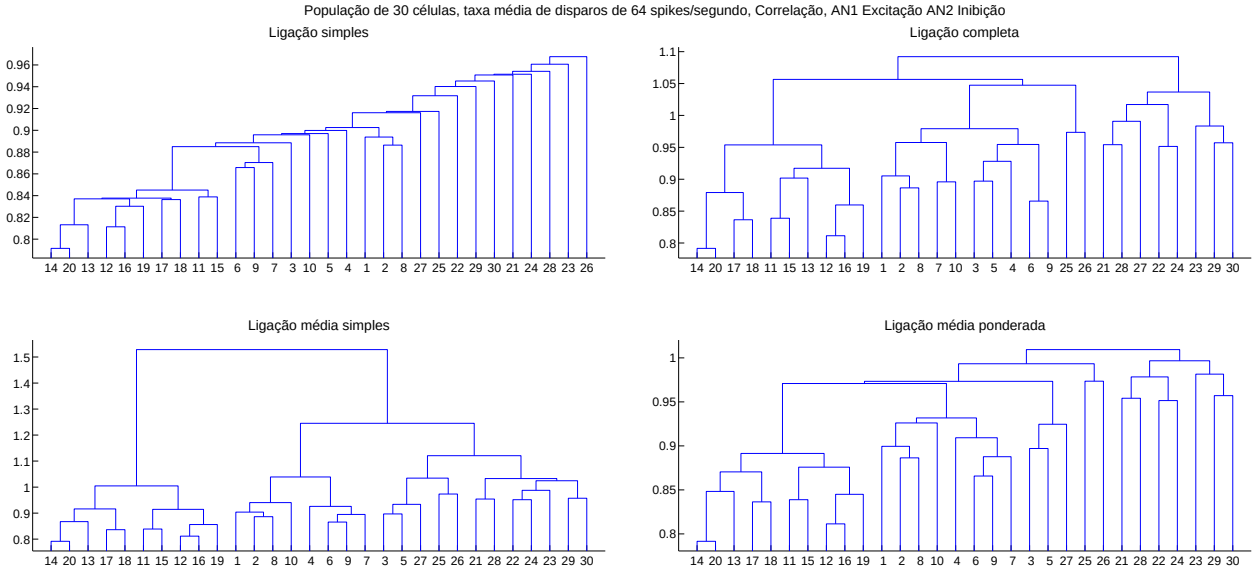


Figura 3.26 – Dendrogramas resultantes do uso da correlação com diferentes tipos de *Linkage* para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Assembleia 1 com interações de excitação e Assembleia 2, de inibição.



distinção visual não é clara.

Observando a ligação simples para as demais métricas de similaridade notamos uma maior facilidade para a identificação da AN com interações inibitórias, que tem folhas mais evidentemente separadas. Já a AN de interações excitatórias, foi corretamente agrupada em uma folha apenas para a correlação e para a distância SPIKE. Para as métricas de Victor e Purpura e de Van Rossum eles não formam folhas e aparecem como os últimos nós do dendrograma. Para a métrica ISI não observamos um agrupamento estruturado para essas células.

Para a ligação média simples, notamos que a distinção das ANs foi clara, porém, era difícil diferenciar se as células eram independentes ou se constituíam de fato uma AN. Ao dividirmos os dendrogramas em 3, sendo uma parte correspondente às células independentes e as demais à cada uma das ANs, notamos que existiam alguns neurônios incorretamente classificados para a AN excitatória nas métricas de similaridade correlação, ISI e distância SPIKE. Observamos também que a altura das folhas das métricas de Victor e Purpura e Van Rossum era maior proporcionalmente ao aumento do número de disparos na interação entre os elementos das ANs.

3.1.4.0.2 Agrupamento Hierárquico: Simulações utilizando trens de *spike* que consideram a correlação entre neurônios em ANs sem neurônios disjuntos

Nesta etapa, analisamos os dendrogramas do conjunto de dados gerado segundo o modelo que utiliza a correlação entre neurônios em uma janela de 10 ms e taxa média de disparos de 22,37 $spikes/segundo$ para uma população de 30 células com ANs independentes, descrita na Seção 3.1.1. Exibimos apenas o cenário com duas ANs que interagem de maneira excitatória, assim como um ponto de interação fraca, em que as células possam ser consideradas aproximadamente independentes entre si.

Analisamos cada uma das formas de *Linkage*, exibidas para a Correlação na Figura 3.31, para um cenário de células independentes e na Figura 3.32 para as interações de excitação.

Observamos que, diferentemente do resultado obtido para o modelo de Poisson, esta abordagem possibilitou a identificação visual das ANs em todas as formas de *Linkage*, porém, permaneceu a dificuldade em identificar células independentes para a ligação média simples.

Para a métrica de similaridade de Hamming, assim como na modelagem de Poisson, obtivemos dendrogramas que não agruparam as ANs corretamente.

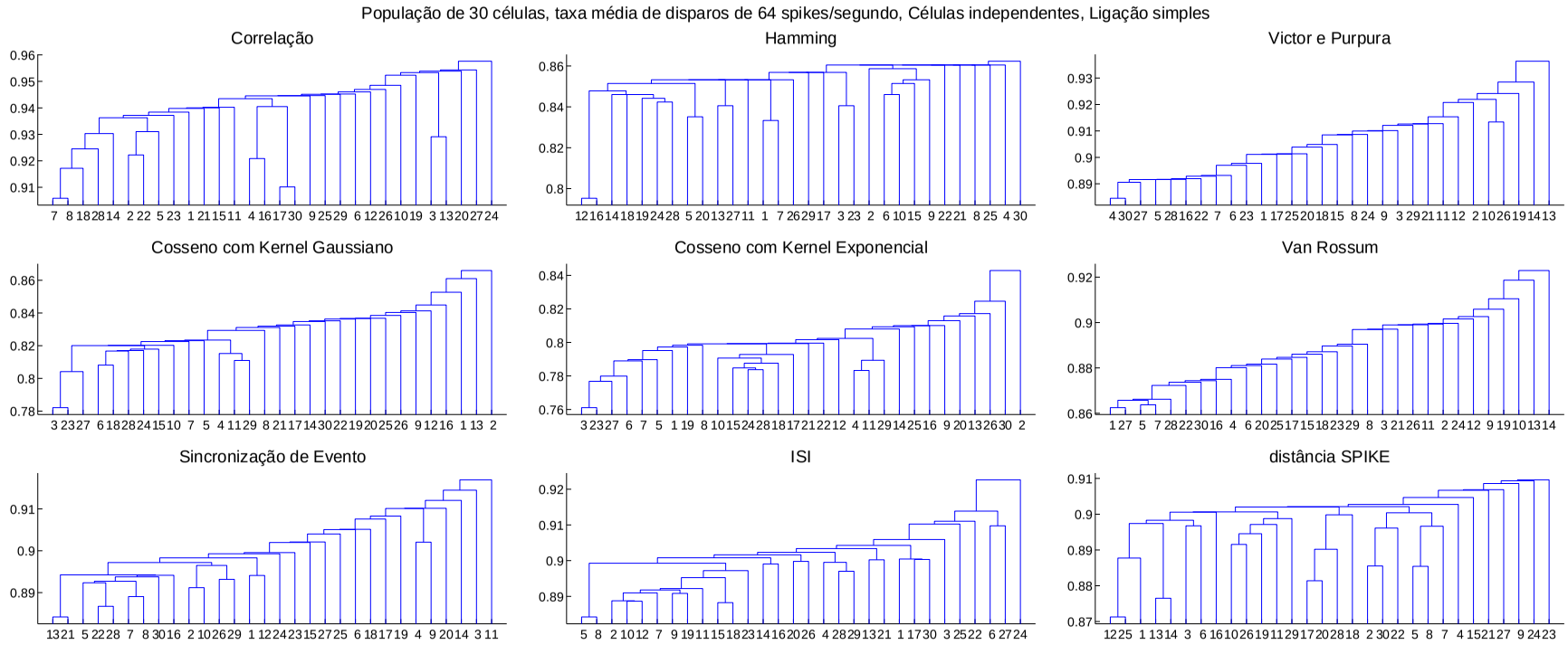


Figura 3.27 – Dendrogramas resultantes do uso de todas as métricas de similaridade com diferentes a ligação simples para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Células Independentes.

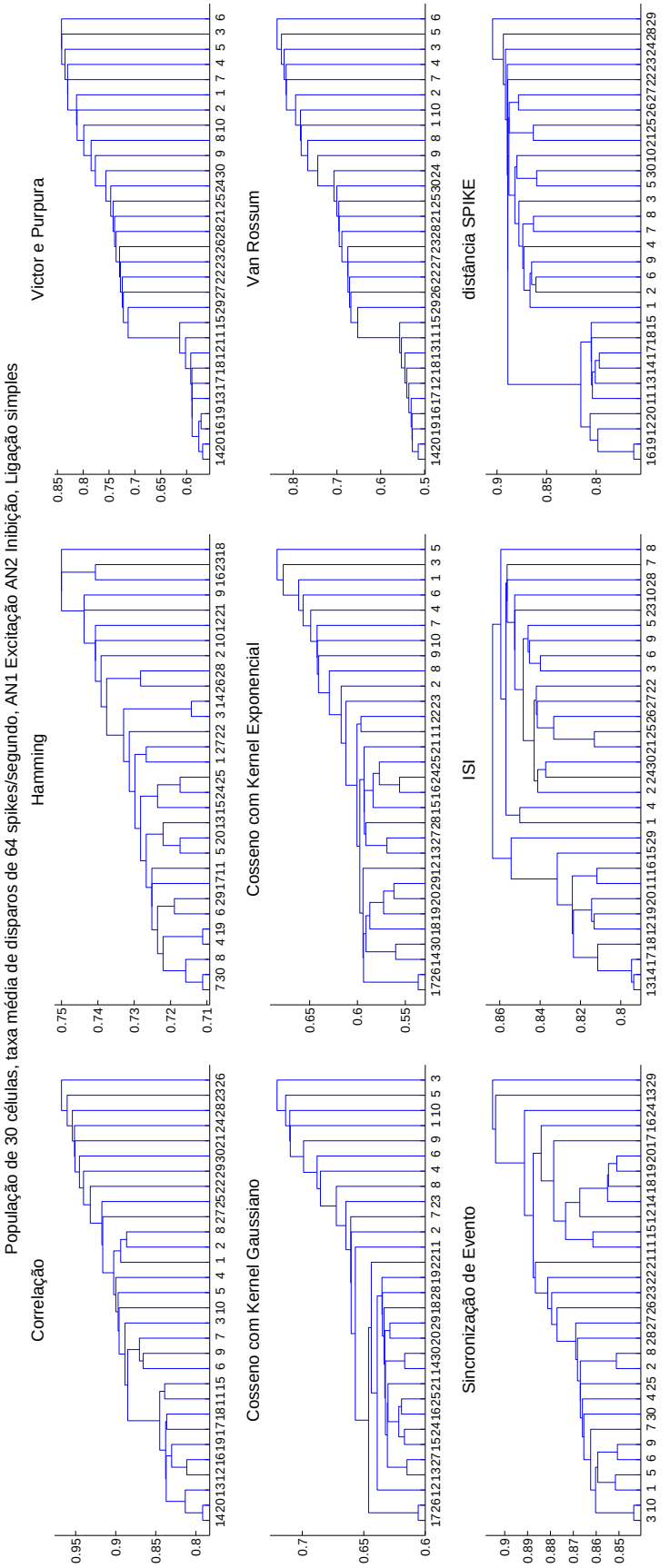


Figura 3.28 – Dendrogramas resultantes do uso de todas as métricas de similaridade com diferentes a ligação simples para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Assembleia 1 com interações de excitação e, assembleia 2, de inibição.

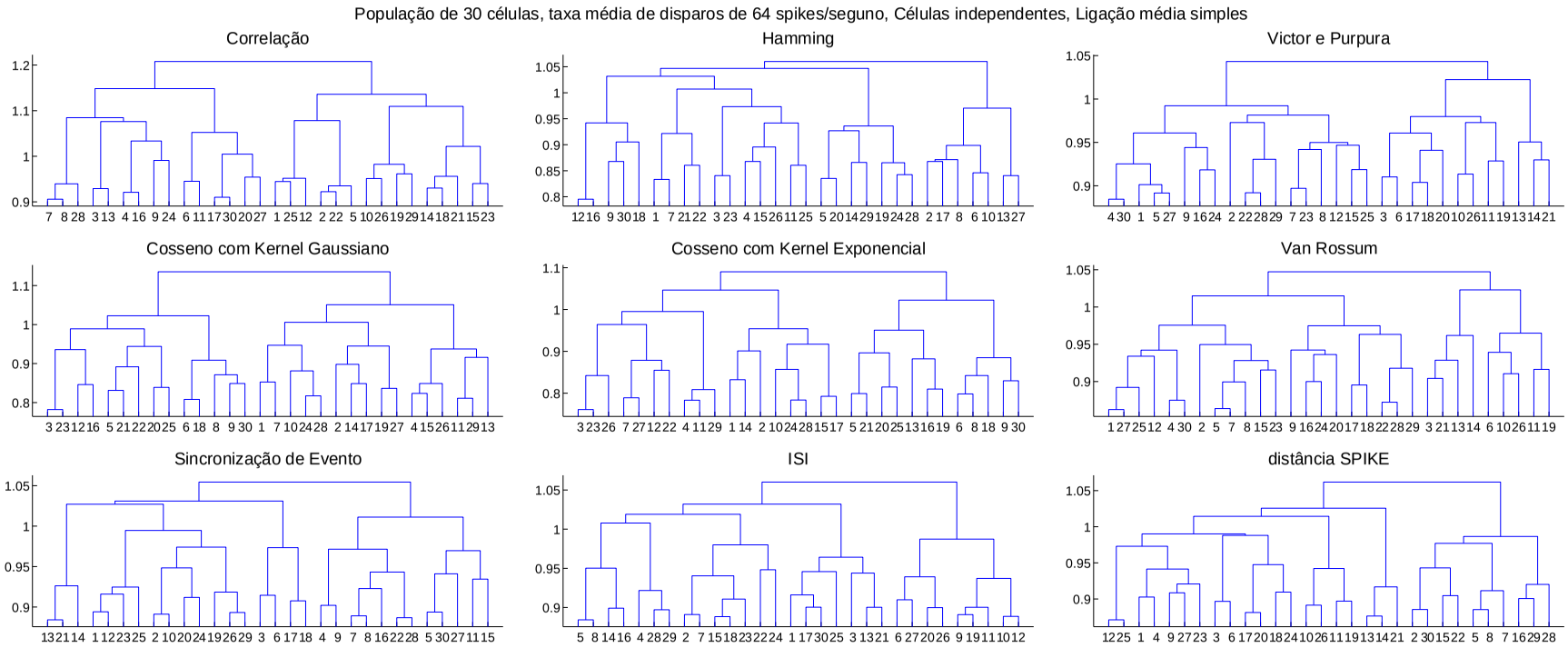


Figura 3.29 – Dendrogramas resultantes do uso de todas as métricas de similaridade com diferentes a ligação média simples para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Células Independentes.

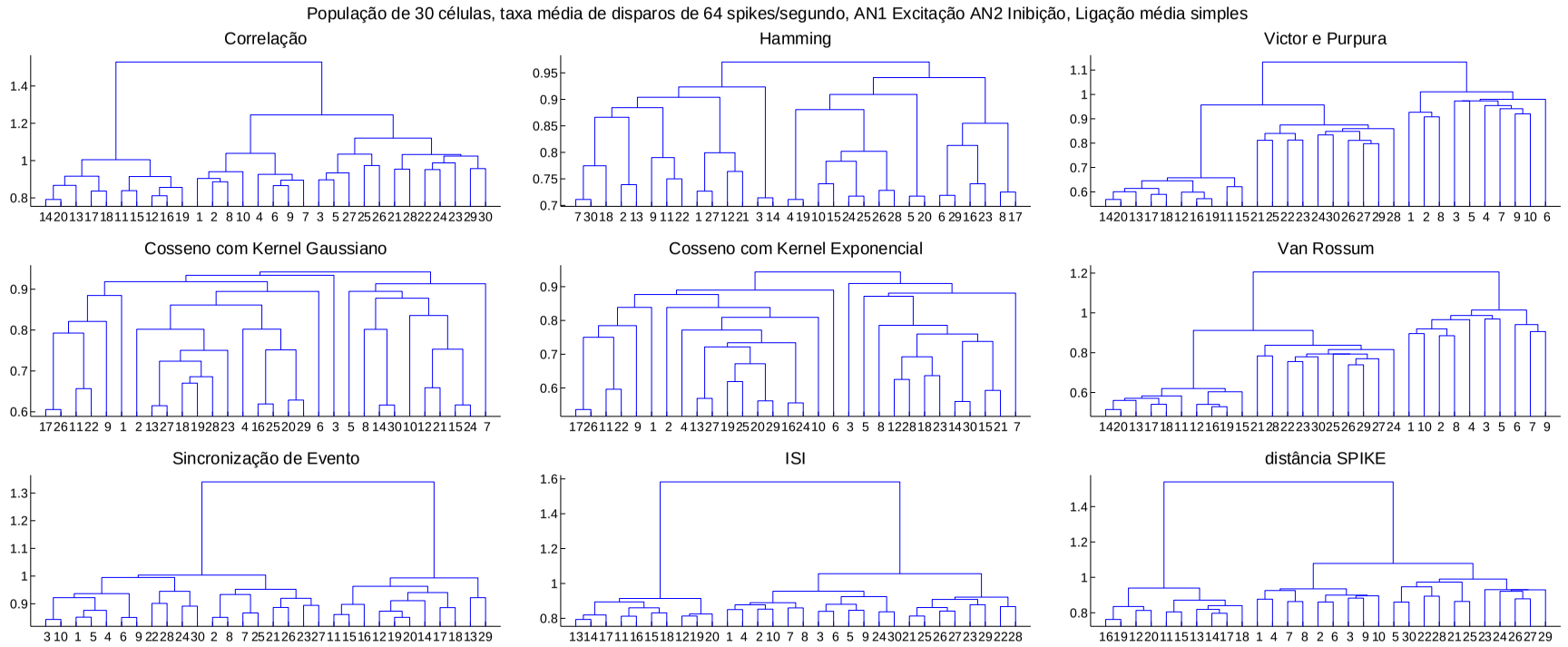


Figura 3.30 – Dendrogramas resultantes do uso de todas as métricas de similaridade com diferentes a ligação média simples para uma população de 30 células com taxa média de disparos de 64 *spikes/segundo* e ANs independentes. Assembleia 1 com interações de excitação e, assembleia 2, de inibição.

Figura 3.31 – Dendrogramas resultantes do uso da correlação com diferentes tipos de *Linkage* para uma população de artificial de 30 células com taxa média de disparos de 22,37 *spikes/segundo* geradas segundo o modelo de Macke *et al.* (2009) e ANs independentes. Células Independentes.

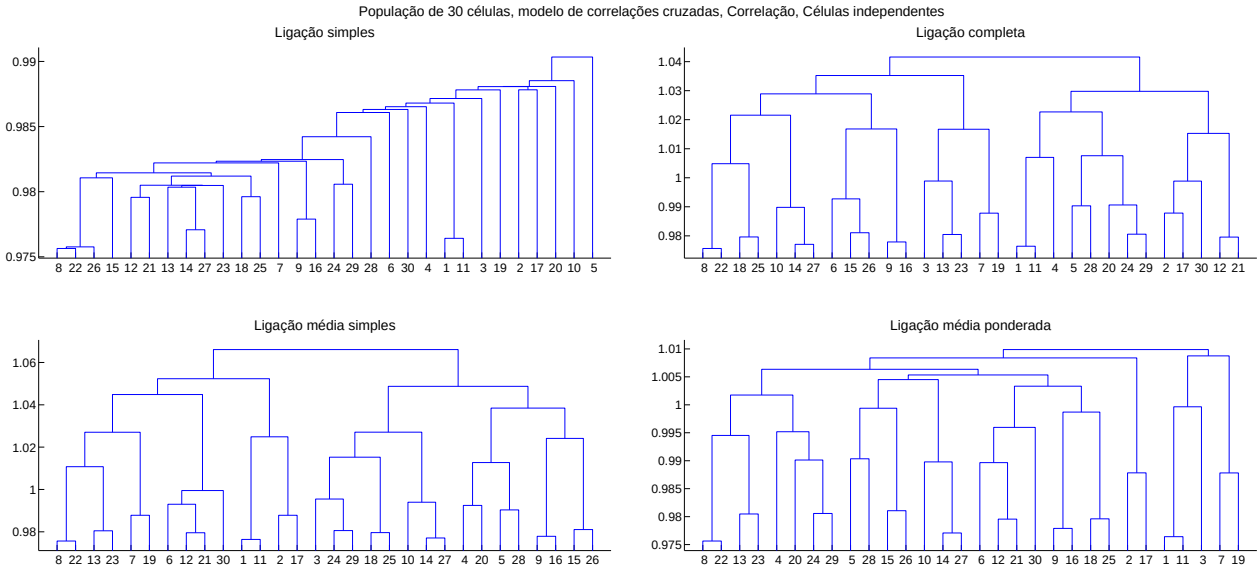
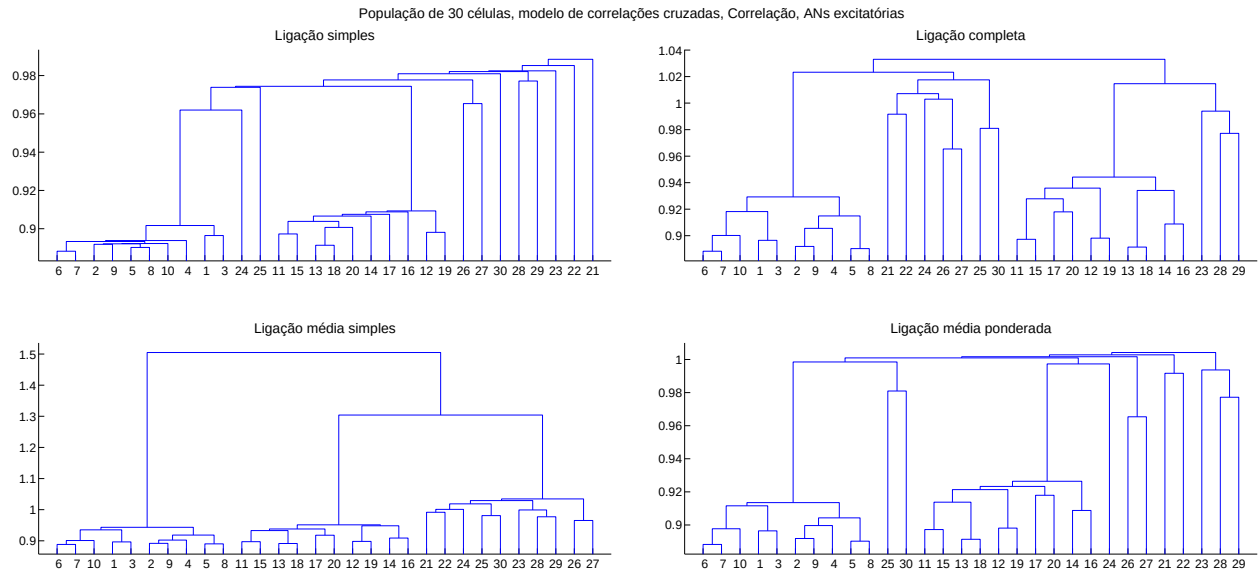


Figura 3.32 – Dendrogramas resultantes do uso da correlação com diferentes tipos de *Linkage* para uma população de artificial de 30 células com taxa média de disparos de 22,37 *spikes/segundo* geradas segundo o modelo de Macke *et al.* (2009) e ANs independentes. Assembleias com interações de excitação.



3.1.5 Resultados: Estabilidade de uma partição

Para finalizar a análise de algoritmos, verificamos o desempenho da Estabilidade de uma partição ao ser associada às diferentes métricas de similaridade. Iniciaremos a análise para ANs independentes e prosseguiremos para cenários em que as ANs compartilham neurônios. Além disso, utilizaremos os conjunto de dados gerados segundo o modelo de Poisson e para cenários em que a população foi descrita por meio da correlação cruzada entre os membros.

3.1.5.1 Estabilidade de uma partição: Simulações de ANs Independentes

Iniciamos com o conjunto gerado segundo a modelagem de Poisson, com 30 células e ANs disjuntas, conforme descrito na seção 3.1.1 com a taxa média de disparos obtida experimentalmente, de 22,37 *spikes/segundo*. Exibimos apenas o resultado para a métrica de similaridade de Billeh, pois foi a única a apresentar resultados não nulos. A Figura 3.33a mostra o desempenho para a detecção de ANs (Equação 2.37) e, a Figura 3.33b, para a classificação dos neurônios (Equação 2.39). Notamos que, neste caso, o particionamento automático implementado retornou uma grande quantidade de falsos positivos na detecção de ANs, o que pôde ser verificado no cenário em que os trens eram independentes entre si (região central da figura). Notamos também que na região de interações excitatórias para as duas ANs houve um aumento no desempenho para a classificação dos neurônios, atingindo o patamar de 50%. Ao avaliarmos a classificação dos neurônios, observamos que sempre que o conjunto de elementos era dividido em comunidades, eles eram alocados a alguma das comunidades. Não observamos elementos que permaneceram independentes em nossas simulações.

A seguir, analisamos o resultado para o conjunto de dados com a mesma topologia, porém, com a taxa média de disparos aumentada para 64 *spikes/segundo*. Mais uma vez, apenas a métrica de Billeh gerou um resultado não nulo para os cenários estudados. O desempenho para a detecção de ANs (Equação 2.37) e classificação dos neurônios (Equação 2.39) seguem nas Figuras 3.34a e 3.34b, respectivamente. Notamos que houve um aumento expressivo no desempenho de detecção de ANs para a região em que as interações entre as células são de excitação (primeiro quadrante da figura). Esse efeito foi acompanhado pelo desempenho na classificação dos neurônios que, no entanto, permaneceu no patamar de 50%. Alguns pontos isolados apresentaram desempenho de 80%, no entanto, nesses casos, foram detectadas 3 ANs e os neurônios independentes foram agrupados em uma única assembleia. Permaneceram os falsos positivos detectados na região central, em que os neurônios são independentes.

A próxima etapa consiste em analisar o desempenho do algoritmo à medida que au-

Figura 3.33 – Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 30 células com ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.

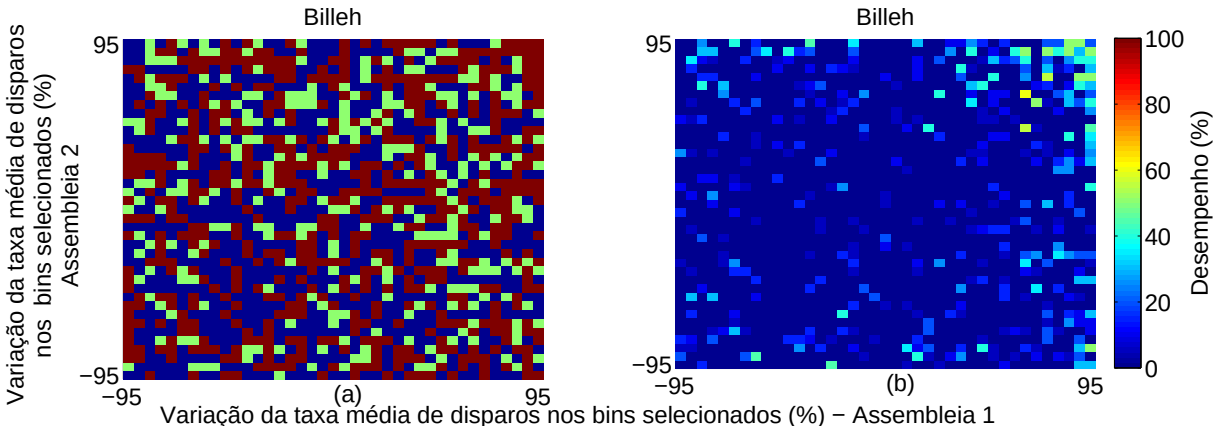


Figura 3.34 – Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 30 células com ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.

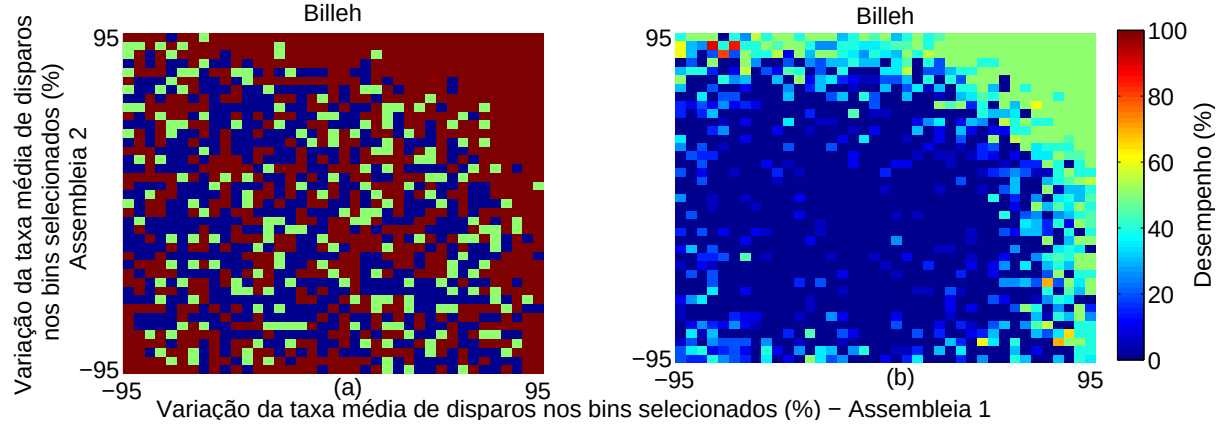
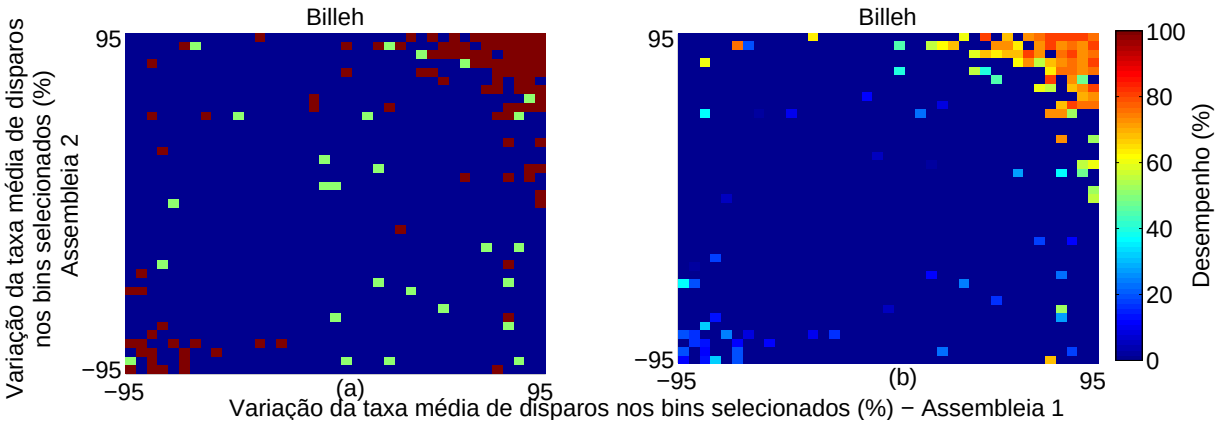


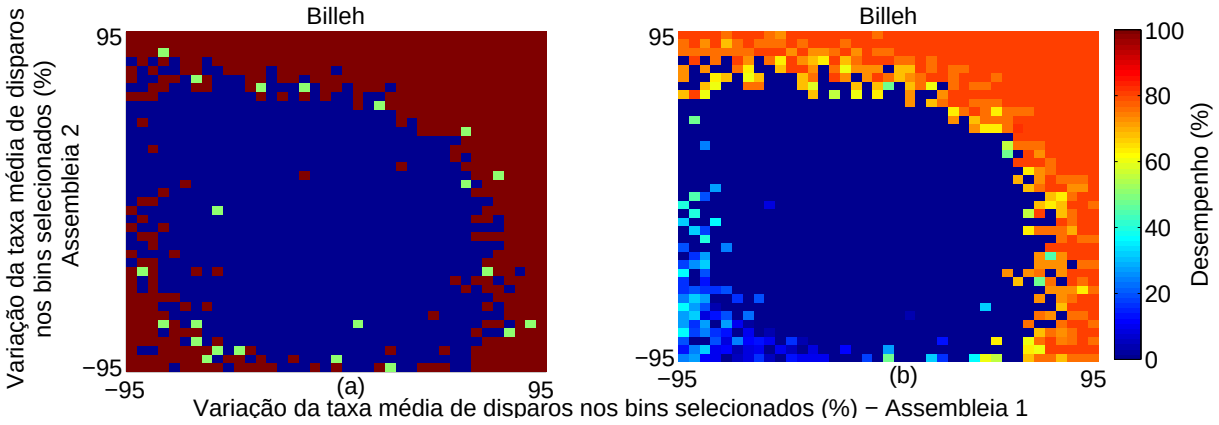
Figura 3.35 – Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 22,37 *spikes/segundo*. Simulamos uma população de 60 células com ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.



mentamos as comunidades. Assim, passamos ao conjunto com 60 células independentes, descrito no caso 2 da Seção 3.1.1, ainda gerado seguindo o modelo de Poisson e com uma taxa média de disparos de 22,37 *spikes/segundo*. Continuamos obtendo resultados não nulos somente para a métrica de similaridade de Billeh. Exibimos na Figura 3.35a o desempenho para a detecção de ANs (Equação 2.37) e, para a classificação dos neurônios (Equação 2.39), na Figura 3.35b. Observamos que o desempenho detecção de ANs se mantém elevado na região em que as duas assembleias tem interações de excitação, porém que os falsos positivos diminuem. Para a classificação dos neurônios, o desempenho máximo obtido foi de 80%.

Dando prosseguimento, analisamos o desempenho para o conjunto de dados que contém ANs maiores com uma taxa média de disparos aumentada, de 64 *spikes/segundo*. Mais uma vez, apenas a métrica de Billeh retornou resultados não nulos. Nas Figuras 3.36a e 3.36b, a seguir, que correspondem ao desempenho para a detecção de ANs (Equação 2.37) e para a classificação dos neurônios (Equação 2.39), respectivamente. Observamos que o desempenho detecção de ANs foi elevado na região em que as ao menos uma das assembleias tem interações de excitação, numa região maior do que no cenário com uma taxa média de disparos menor (Figura 3.35a). Ainda, a baixa ocorrência de falsos positivos se mantém. Além disso, o desempenho aumentou também na região que as duas ANs possuíam interações de inibição (terceiro quadrante da figura). Para a classificação dos neurônios, o desempenho máximo obtido continuou em 80 %. Observamos que a classificação apresentou um desempenho inferior

Figura 3.36 – Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células com ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.



para o caso que as duas ANs tinham interações de inibição.

A seguir, analisamos a resposta do algoritmo a atrasos de fase entre membros de uma AN. Continuamos com a população de 60 células, com ANs maiores e independentes e com uma taxa média de disparos de 64 *spikes/segundo*, já que este cenário torna a visualização dos efeitos mais clara. A interação de 5 neurônios foi atrasada em um *bin*. O resultado pode ser verificado para a métrica de Billeh na Figura 3.37a, para a detecção de ANs (Equação 2.37) e, na Figura 3.37b, para a classificação dos neurônios (Equação 2.39). Comparativamente ao cenário sem atrasos (Figura 3.36) observamos uma queda no desempenho para a classificação dos neurônios na região de interações excitatórias da assembleia 2.

Na próxima etapa, analisamos os resultados da aplicação do algoritmo ao conjunto de dados gerado segundo o modelo que utiliza a correlação entre neurônios em uma janela de 10 *ms* e taxa média de disparos de 22,37 *spikes/segundo* para uma população de 30 células com ANs independentes, descrita na Seção 3.1.1. Tanto para a correlação quanto para a métrica de Hamming obtivemos um desempenho nulo em todos os cenários analisados.

3.1.5.2 Estabilidade de uma partição: Simulações de ANs que compartilham neurônios

Assim como a Modularidade, a Estabilidade de uma partição também possui limitações para lidar com ANs que compartilham neurônios, como será discutido na Seção 4.1.2.

Figura 3.37 – Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células com ANs independentes. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.A atividade de 5 neurônios de uma das ANs foi atrasada um *bin*.

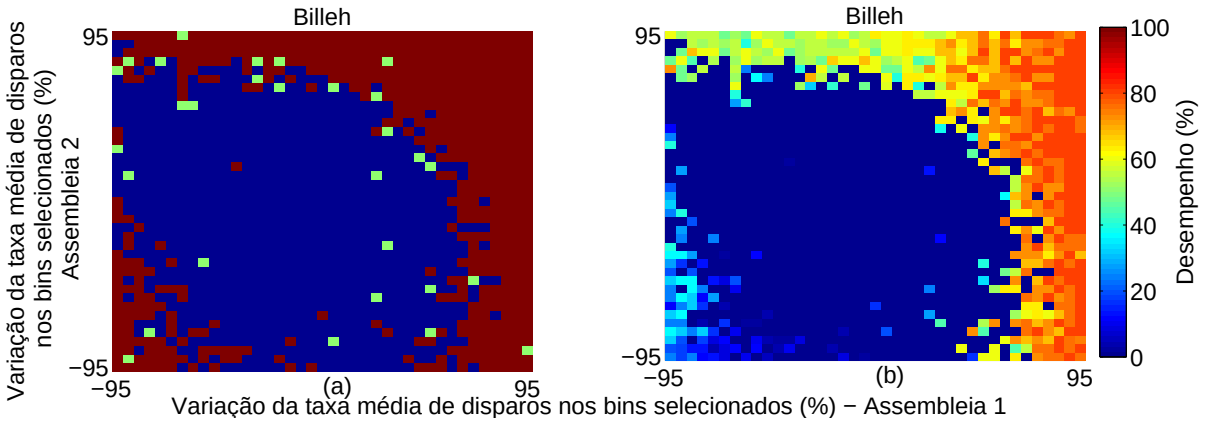
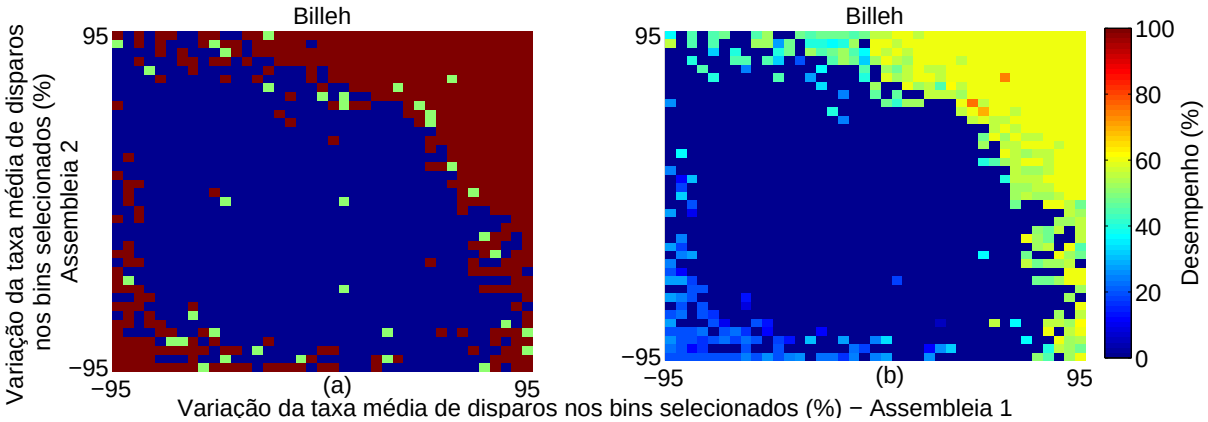


Figura 3.38 – Desempenho da Estabilidade de Markov associada à métricas de similaridade de Billeh na detecção do número de ANs (Seção 2.4.1), na primeira figura, e, na classificação dos neurônios, na segunda (Seção 2.4.2), quando aplicado a trens artificiais gerados segundo o modelo de Poisson com uma taxa de disparos basal de 64 *spikes/segundo*. Simulamos uma população de 60 células com ANs com 5 neurônios em comum. Em um total de 10^4 amostras e selecionamos 300 *bins* aleatoriamente em que a taxa de disparos do modelo de Poisson foi escalonada por um fator.



Para ilustrar esse efeito, exibimos na Figura 3.38 o desempenho do algoritmo aplicado ao conjunto de dados composto por ANs maiores (caso 2, descrito na Seção 3.1.1) e com taxa média de disparos de 64 *spikes*/*segundo*, pois, neste cenário, as diferenças entre os cenários foi mais evidente. Ao compararmos o desempenho com o de ANs sem neurônios em comum (Figura 3.38), observamos uma queda para a detecção do número de ANs no segundo e no quarto quadrantes, quando a interação entre os elementos de uma das assembleias se torna mais inibitória. Principalmente, notamos que a classificação dos neurônios deixou de atingir o patamar de 80% e passou a saturar em 60%. Dois casos isolados exibiram desempenhos maiores, de 74%, mas, nesses casos, três ANs foram detectadas.

3.2 Resultados da aplicação dos algoritmos a registros eletrofisiológicos extracelulares

A partir dos resultados das simulações exibidas ao longo da Seção 3.1, selecionamos algumas metodologias para a caracterização das ANs e aplicamos a registros eletrofisiológicos extracelulares de seis ratos *Long-Evans* engajados na tarefa de discriminação tátil, descrita na Seção 2.5. As metodologias selecionadas foram o ICA associado à Correlação e ao ISI. Exemplificaremos também limitações do uso da Modularidade, da forma descrita neste trabalho, em conjunto com a métrica Cosseno e da Estabilidade de uma partição associada com a métrica de Billeh.

Adotamos como referência o instante que demarcava a discriminação tátil. Extraímos a atividade das células em janelas de três segundos centradas neste evento e concatenamos os segmentos selecionados para cada uma das repetições do experimento. Caracterizamos o conjunto de dados de cada um dos animais na Tabela 3.1, com informações sobre o número de células registradas de cada uma das áreas, o número de repetições do experimento e a taxa média de disparos por segundo.

Tabela 3.1 – Caracterização dos dados experimentais analisados.

	Rato 1	Rato 2	Rato 3	Rato 4	Rato 5	Rato 6
Número de células PFC	23	17	27	16	23	20
Número de células PPC	16	17	22	20	10	9
Número de células S1	14	8	12	9	19	15
Número de células V1	26	23	20	21	43	38
Número de repetições do experimento	80	110	198	249	216	196
Taxa média de disparos (*spikes*/s)	9,79	13,28	17,39	16,70	16,92	22,41

A primeira etapa foi a análise das PCs correspondentes a cada uma das assembleias, para que o ganho G, descrito na Equação 2.25 fosse definido. Na Figura 3.39 exibimos as PCs correspondentes a cada uma das ANs do Rato 1, resultantes da aplicação da Correlação associada ao ICA. Nesta análise utilizamos $G = 1$, porém adicionamos um termo extra correspondente a um desvio padrão de $|w_k|$, para $k = \{1, ...n_{Neurons}\}$, para melhorar a distinção entre as células da AN e as independentes. O valor de Thr resultante foi demarcado em vermelho. Observamos que a distribuição dos pesos w_k, $k = \{1, ...n_{Neurons}\}$ de algumas PCs apresentaram desvios no padrão característico. Como ilustrado nas Figuras 2.6a e 2.6c, quando uma assembleia segue o padrão característico seus pesos tem valores parecidos enquanto os outros neurônios tem pesos nulos. A Figura 3.39a exibe um exemplo em que os pesos tem valores diferentes e as Figuras 3.39b e 3.39c exemplificam casos em que é difícil encontrar um limiar que diferencie um nerônio independente dos neurônios que compõem as ANs.

Exibimos o resultado do ICA associado à Correlação e ao ISI para o Rato 1 na Figura 3.40 a seguir, onde cada pequeno círculo correspondia a uma das células registradas e, cada cor representava uma das áreas, como descrito pela legenda. As linhas, por sua vez, conectavam as células que participavam das ANs detectadas. Observamos uma correspondência entre as assembleias identificadas por meio de cada um dos métodos. Para o Rato 1, notamos que a AN1 da Figura 3.40a tem correspondência com as assembleias 2 e 4 da Figura 3.40b. Já as assembleias 2 e 3 da Figura 3.40a, correspondem às assembleias 1 e 3 da Figura 3.40b, respectivamente.

Além disso, a análise do histograma centrado no evento possibilitou visualizarmos que as células selecionadas para comporem as ANs apresentam padrões de modulação similares. Exibimos na Figura 3.41 o histograma das células do córtex somatossensorial primário do Rato 1. O intervalo de referência anterior à discriminação adotado foi de -1,5 a -0,5 segundos e foi demarcado em magenta; uma linha pontilhada em magenta correspondia ao valor médio da taxa de disparos neste segmento. A demarcação amarela correspondia ao intervalo de tempo em que comparamos a atividade da célula com aquela obtida durante o período de referência. Em azul, indicamos os instantes em que houve uma resposta significativa de inibição e, em vermelho, de excitação. Os limiares de significância foram definidos utilizando uma técnica de soma cumulativa descrita por Wiest *et al.* (2010). Como pode ser observado no histograma das células #1, #2, #3, #4 e #8, que compõe a AN, a taxa média de disparos destas células decai significativamente após a discriminação. Já para a célula #6, houve um aumento significativo da taxa média de disparos após o evento. Foi possível observar também que a atividade de algumas células não parece ter ligação direta com o evento, como as células #10 e #11, e que outras células, como a #9, a #12, a #13 e a #14 parecem seguir outras formas de modulação.

Figura 3.39 – PCs resultantes da aplicação da correlação associada ao ICA nos registros eletrofisiológicos extracelulares do Rato 1. Os neurônios k, para $k = \{1, ...n_{Neurons}\}$, cujos pesos w_k foram maiores que o limiar (em vermelho) foram considerados como membros da assembleia neural.

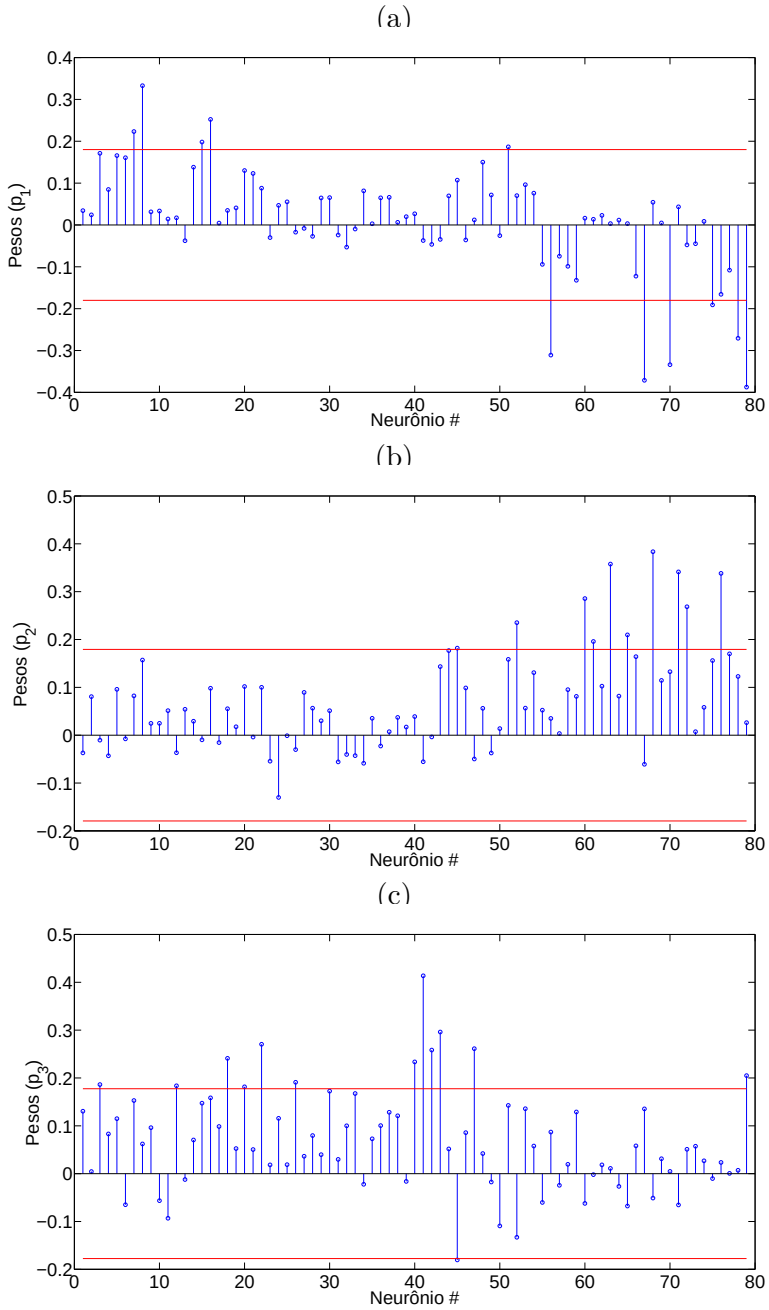


Figura 3.40 – Resultado da aplicação do ICA associado à correlação (Figura 3.40a) e ao ISI (Figura 3.40b) para o Rato 1. Para cada repetição do experimento, segmentos de dados foram selecionados ao redor da discriminação tátil.

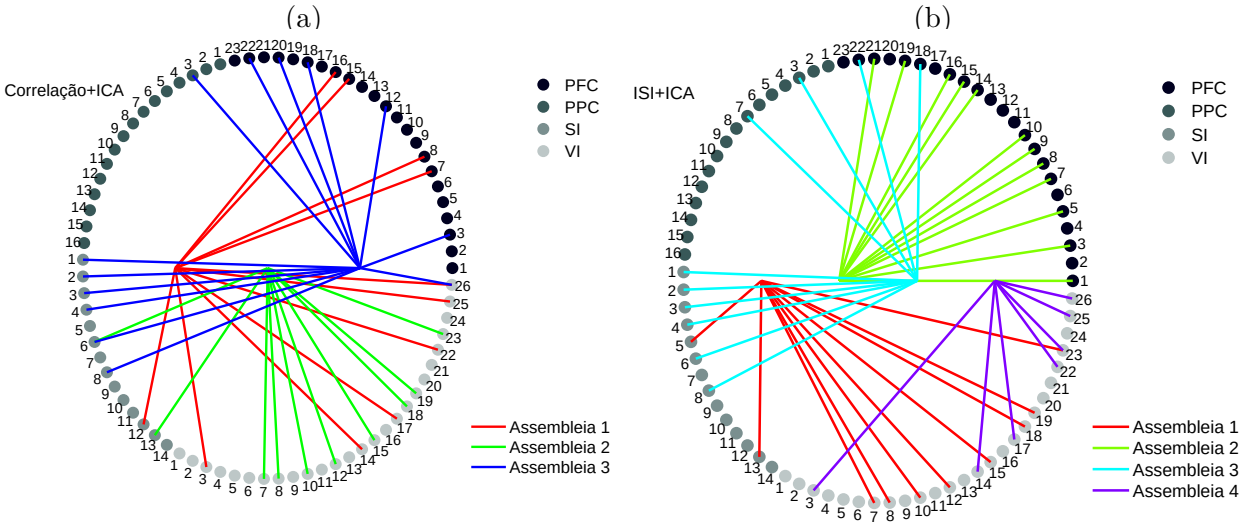


Tabela 3.2 – Correspondência entre o uso das métricas de similaridade correlação e ISI associadas ao ICA para o conjunto de dados dos animais 1,2 e 3.

	Rato 1			Rato 2				Rato 3			
ICA+Correlação	AN1	AN2	AN3	AN1	AN2	AN3	AN4	AN1	AN2	AN3	AN4
ICA+ISI	AN2+AN4	AN1	AN3	AN4	AN2	AN1	AN3	AN4	AN2	AN1	AN3
Células em comum	10	9	9	11	11	9	6	15	8	18	14
Células distintas (I-C)	9	1	1	0	0	0	0	0	0	0	0
Células distintas (C-I)	1	1	4	0	0	0	0	0	0	0	0

Os animais analisados apresentaram uma consistência entre as ANs detectadas ao aplicarmos o ICA com as métricas de correlação e ISI, i.e. obteve-se o mesmo resultado com ambos os métodos, como pode ser verificado nas Tabelas 3.2 e 3.3, a seguir. Nas duas primeiras linhas, mostramos a correspondência das ANs encontradas. Na terceira, contabilizamos o total de células identificadas simultaneamente pelas duas metodologias. Na quarta, contabilizamos o número de células que foram identificadas pelo ISI e não foram identificadas pela correlação (I-C). De maneira complementar, na quinta, contabilizamos o número de células encontradas pela correlação que não são detectadas pelo ISI (C-I). Notamos que para os Ratos 2, 3 e 5 a correspondência entre os métodos é perfeita e que para os animais 1, 4 e 6 algumas células são distintas ou as ANs não apresentam correspondência.

Na próxima etapa, observamos em quais áreas corticais se localizavam as células que participavam de cada uma das ANs. O resultado obtido para a aplicação da Correlação em conjunto com o ICA foi descrito nas Tabelas 3.4 e 3.5, a seguir. Notamos que as interações

Figura 3.41 – Histograma centrado no evento de discriminação tátil para as 14 células do córtex somatossensorial primário do Rato 1. Observamos que há modulação similar das células que tanto a métrica de correlação como a métrica ISI associadas ao ICA indicaram pertencer a uma mesma AN.

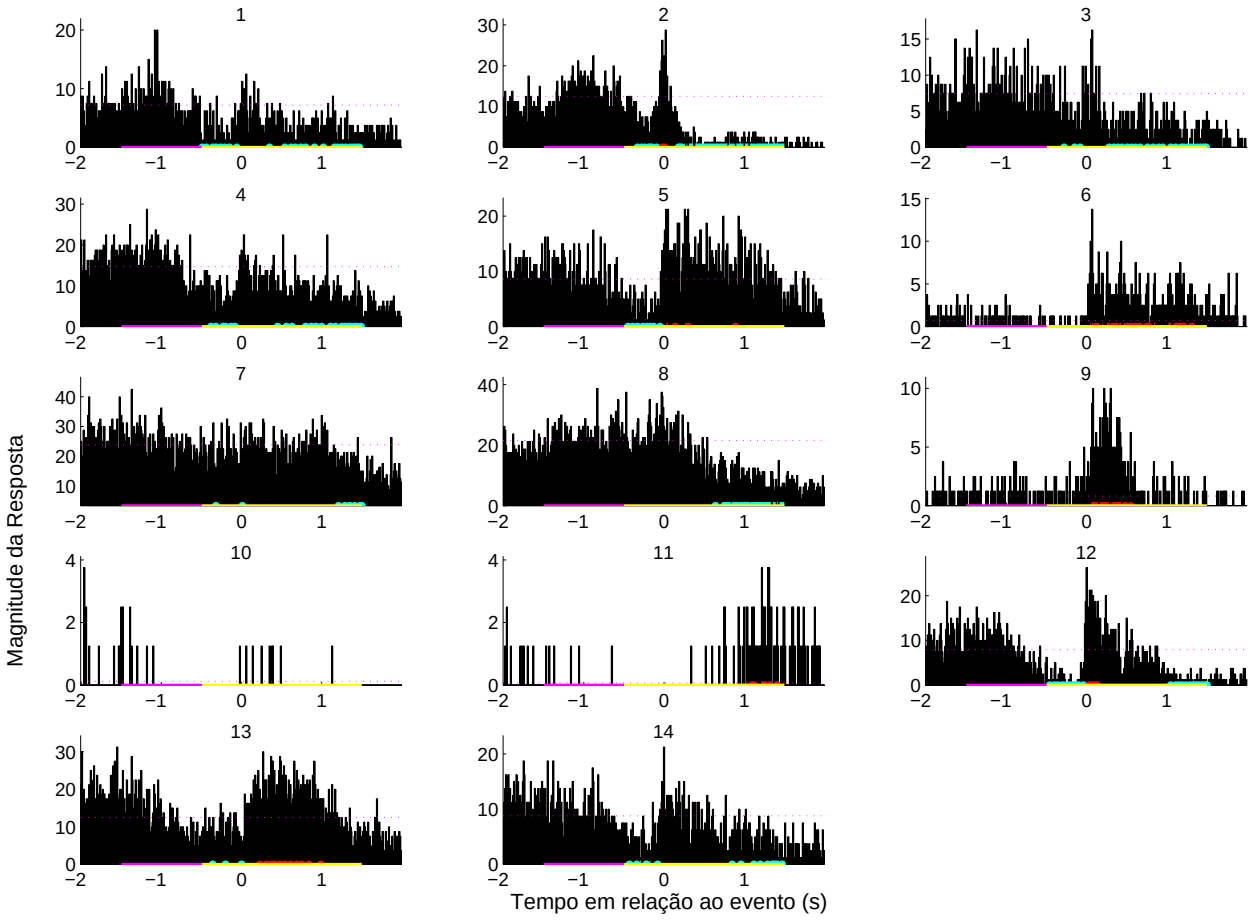


Tabela 3.3 – Correspondência entre o uso das métricas de similaridade correlação e ISI associadas ao ICA para o conjunto de dados dos animais 4,5 e 6.

	Rato 4				Rato 5				Rato 6				
ICA+Correlação	AN1	AN2	AN3	AN4	AN1	AN2	AN3	AN4	AN1	AN2	AN3	AN4	AN5
ICA+ISI	AN3	AN1	-	AN2	AN4	AN1	AN3	AN2	AN2	AN3	-	AN1	-
Células em comum	11	7	-	9	11	15	10	11	9	9	-	8	-
Células distintas (I-C)	1	0	-	0	0	0	0	0	3	5	-	5	-
Células distintas (C-I)	0	0	-	2	0	0	0	0	1	4	-	5	-

entre regiões corticais não foram as mesmas entre todos os animais. Por exemplo, observamos que para os Ratos 1 e 2 poucas associações foram encontradas com células do córtex PPC, enquanto que, para o Rato 3, duas assembleias contavam com mais de três células nestas regiões. Apesar disso, notamos que todos os animais apresentaram formações de ANs em torno do evento de discriminação tátil.

Tabela 3.4 – Áreas que possuem células engajadas na formação de cada uma das ANs. Resultado para a aplicação da Correlação associada ao ICA para os animais 1, 2 e 3.

Assembleias/ Animal	Rato 1	Rato 2	Rato 3
AN1	PFC-V1-S1*	V1	PPC-PFC-V1
AN2	VI-S1*	PFC-PPC*-S1*-V1*	V1
AN3	PFC-S1-PPC*	V1-S1*-PFC*-PPC*	PFC-V1-S1*
AN4	-	V1-S1	PFC-PPC-S1-V1*

* Áreas com menos de 3 células engajadas nesta AN.

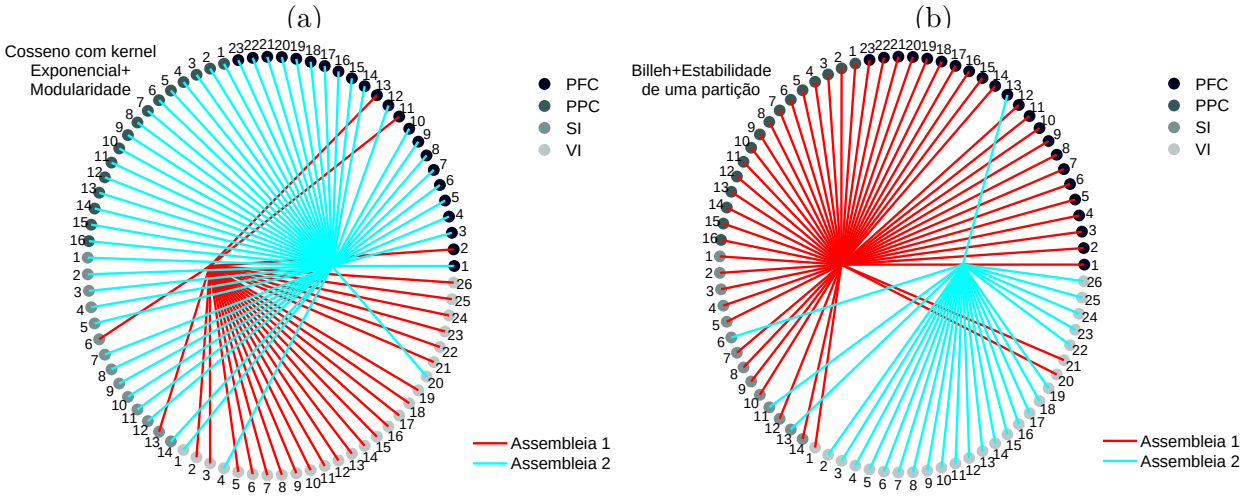
Tabela 3.5 – Áreas que possuem células engajadas na formação de cada uma das ANs. Resultado para a aplicação da Correlação associada ao ICA para os animais 4,5 e 6.

Assembleias/ Animal	Rato 4	Rato 5	Rato 6
AN1	PFC*-PPC*-S1*V1*	PFC-S1-V1	S1-V1-PFC*
AN2	PPC-S1*-V1*	PFC-PPC-S1*-V1*	V1-S1*
AN3	V1	S1-V1-PFC*	PFC-V1
AN4	PFC-V1	-	PFC-PPC-V1-S1*
AN5	-	-	PFC-S1-PPC*-V1*

* Áreas com menos de 3 células engajadas nesta AN.

Para finalizar, exibimos o resultado de aplicação da métrica Cosseno associada ao algoritmo Modularidade (Figura 3.42a) e da Estabilidade de uma partição com a métrica de Billeh (Figura 3.42b) para o Rato 1. É possível observar que o algoritmo apresenta uma limitação pois todas as células são associadas a alguma das comunidades detectadas e, portanto, não existem células independentes.

Figura 3.42 – Resultado da aplicação da Modularidade associada ao cosseno com um *kernel* exponencial (Figura 3.42a) e a métrica de Billeh associada à Estabilidade de uma partição (Figura 3.42b) para o Rato 1. Para cada repetição do experimento, segmentos de dados foram selecionados ao redor da discriminação tátil. É possível notar que células independentes ou que pertençam simultaneamente às duas comunidades não são verificadas.



4 Discussão

O objetivo deste trabalho é o de comparar o desempenho de combinações de métricas de similaridade e métodos de classificação na identificação e caracterização de ANs, destacando limitações, extensões pontuais e cenários mais apropriados para aplicação. Inicialmente, na Seção 4.1, apresentamos o resultado de um estudo das métricas de similaridade descritas na Seção 2.2 e dos algoritmos descritos na Seção 2.3, caracterizando algumas limitações dessas metodologias para o problema de detecção de ANs.

A seguir, discutimos os resultados exibidos no Capítulo 3. Analisamos o desempenho dos métodos de processamento segundo os diferentes parâmetros: taxa média de disparos, tamanho das ANs, categorias de ANs (Figura 2.1), atrasos de fase e modelo utilizado para a geração dos trens de *spike*. Finalizamos a seção com a Tabela 4.1 que resume os achados experimentais encontrados.

Finalmente, comentaremos os resultados de aplicarmos algumas metodologias em um conjunto de registros eletrofisiológicos extracelulares de ratos *Long Evans* realizando uma tarefa de discriminação tátil (Capítulo 2.5).

4.1 Propriedades e limitações das métricas e algoritmos estudados

Após análises qualitativas e simulações, observamos algumas limitações e vantagens tanto das métricas de similaridade quanto dos algoritmos apresentados para caracterizar a atividade de ANs. Nesta seção comentaremos os achados encontrados.

4.1.1 Métricas de similaridade

Inicialmente, fizemos um levantamento de algumas propriedades e limitações relativas às métricas propostas para medir a similaridade entre trens de *spike*, listadas a seguir:

- Correlação: O uso da correlação mostra-se adequado quando a resposta de um conjunto de células modula sua atividade por meio de modificações na taxa média de disparos. Isso pode ser verificado, por exemplo, no trabalho de Sherman (2006), pelo disparo em modo tônico em células de retransmissão do núcleo geniculado lateral em gatos. O operador apresenta uma limitação no tratamento de interações de ordem superior, pois é possível a existência de variáveis com dependência estatística mas correlação nula (PAPOULIS, 1991). Além disso, a métrica não é robusta a atrasos de fase entre

as células (DEOLINDO *et al.*, 2014), como será discutido na Seção 4.2.4 e como pode ser visualizado na comparação da Figura 3.5 e 3.6. Ademais, sendo aplicada à taxa de disparos dos neurônios, ela ignora outras formas de codificação da informação que levam em conta as relações temporais entre os disparos. Finalmente, a computação da correlação é bastante dependente do valor do *bin* escolhido e da duração total dos dados a serem analisados.

- Hamming: Esta métrica pode capturar quando o aumento ou diminuição da atividade de uma célula está intimamente ligada à ocorrência de um evento, como, por exemplo, a situação mostrada por Sherman (2006), em que eventos podem ser detectados por disparos em modo *burst* de células de retransmissão do núcleo geniculado lateral em gatos. No entanto, é importante ressaltar que a métrica possui limitações, pois, no caso de trens muito esparsos, por exemplo, não é possível diferenciar uma interação real de um conjunto aleatório de células que disparam pouco. O mesmo ocorre para situações em que a taxa média de disparos é elevada e, nessas condições, outras métricas de similaridade devem ser utilizadas. Ela também ignora formas mais complexas de interações, além de descartar a informação contida em variações na taxa de disparo.

- Victor e Purpura: Os valores retornados por esta métrica são muito maiores dos que os obtidos pelas matrizes de similaridade originalmente empregadas pelos métodos de classificação, majoritariamente no intervalo $[0, 1]$, em módulo. Para equalizar a métrica com relação às demais, optamos por normalizar a matriz de similaridade segundo o procedimento descrito na Seção 2.2.4, mas outras formas de correção já foram propostas, por exemplo, a descrita por Paiva *et al.* (2010). A utilização da métrica de Victor e Purpura requer o uso de implementações dinâmicas, a fim de que o tempo de processamento não inviabilize sua utilização, já que seu custo computacional é proporcional ao número de *spikes* nos trens a serem comparados (PAIVA *et al.*, 2010). Além disso, a métrica é dependente de um parâmetro variável que define o custo de deslocar as ocorrências dos *spikes* (c_v), definido pelo experimentador, o que, por sua vez, influencia o resultado final.

- Convolução com uma janela e Cosseno: O resultado da convolução de um trem de *spikes* com um *kernel* é dependente de sua escala temporal, definida por sua largura (σ). *Kernels* com larguras pequenas selecionam características referentes à sincronia dos disparos, e, larguras grandes, à taxa de disparos. Portanto, a escolha de σ tem bastante influência no resultado final. O uso de diferentes *kernels* foi avaliado neste trabalho e foi possível verificar que, para a escala temporal utilizada, nossos resultados estão em concordância com Paiva *et al.* (2010), pois notamos poucas diferenças entre o uso de uma

função de *kernel* em relação a outra, como pode ser verificado, por exemplo, nas Figuras 3.4, 3.17, 3.19, 3.20 e 3.21. Especificamente para a Modularidade, a associação com o cosseno em taxas médias de disparos menores resultou em diferenças mais significativas entre os *kernels* (Figuras 3.16 e 3.18), porém, o resultado estava muito sujeito a falsos positivos. Além disso, já foi reportado por Fellous *et al.* (2004) uma limitação quando a dinâmica da atividade populacional submetida à métrica retorna valores numa faixa muito estreita de amplitude, e que pode ser necessário um reescalonamento do resultado a fim de aumentar a eficácia dos métodos de classificação.

- Van Rossum: A convolução possibilita à métrica englobar atrasos temporais entre disparos até uma escala determinada pela magnitude de σ e só assumirá o valor mínimo - zero- no caso de trens completamente idênticos. Ela possui uma limitação em termos computacionais, uma vez que, apesar de otimizações já propostas (PAIVA *et al.*, 2010), sua complexidade impossibilita o processamento de conjuntos extensos de dados. Neste trabalho foi possível verificar que as métricas levam a resultados similares entre si, corroborando com as observações de Houghton e Victor (2010). Isso pode ser observado, por exemplo, ao compararmos as Figuras 3.5c e 3.5i além dos dendrogramas das Figuras 3.28 e 3.30.

- Billeh: A métrica proposta por Billeh *et al.* (2014) incorpora interações do tipo direcionais, sendo a única dentre as analisadas neste trabalho a operar com matrizes potencialmente assimétricas. Isso impossibilita sua utilização em conjunto com os algoritmos explorados neste trabalho distintos ao qual ela foi originalmente associada (estabilidade de uma partição). Conforme mencionado na Seção 2.3.1, o paralelo entre o SVD e o PCA só é adequado quando os autovetores são independentes, condição que requer matrizes simétricas. Já a proposta original para a modularidade de Newman (2006), descrita na Seção 2.3.3, considera apenas redes não direcionadas, o que mais uma vez restringe a aplicação a matrizes simétricas. Cabe ressaltar que algumas melhorias já foram propostas para esse algoritmo para englobar interações direcionais, como a proposta de Leicht *et al.* (2007), e a análise do resultado empregando essa métrica se tornaria possível. Finalmente, o uso do Agrupamento Hierárquico não é apropriado quando o requisito de simetria da matriz de similaridade é violado conforme já mencionado na Seção 2.3.4. Além disso, apesar de a métrica conseguir distinguir interações de excitação e inibição, o que, até onde podemos verificar, é pouco explorado na literatura, ela não consegue classificar a atividade de uma célula nessas categorias de maneira automática. Esta informação deve ser fornecida de maneira complementar pelo experimentador, o que poderia inviabilizar sua usabilidade, já que a tendência é que o número de células registradas cresça com as contínuas melhorias em técnicas de registro

eletrofisiológico extracelular (NICOLELIS *et al.*, 1997; BUZSáKI, 2010; SCHWARZ *et al.*, 2014).

- Sincronização de evento: A definição do parâmetro τ de forma que ele se adapte dinamicamente ao longo do tempo segundo variações instantâneas na taxa de disparos é interessante, pois torna-a independente da ação do experimentador. No entanto, isso eleva consideravelmente o tempo de processamento da matriz de similaridade para conjuntos de dados com longa duração e exige a paralelização do processamento à medida que o número de células analisadas aumenta.

- Intervalo entre spikes (ISI): Esta métrica também é não paramétrica, com uma escala temporal intrínseca e adaptativa. Suas limitações para distinguir situações em que a taxa de disparos instantânea é constante ao longo do tempo foram levantadas na Seção 2.2.3.2. De maneira análoga à métrica de Sincronização de evento, a computação da matriz de similaridade S tem um elevado custo computacional.

- Distância SPIKE: A métrica corrige o problema presente na métrica ISI, levantado no item anterior. No entanto, conforme mencionado na Seção 2.2.3.3 ela é uma *semi-distância* e pode desobedecer a desigualdade triangular. Seu custo computacional para o cálculo da matriz de similaridade S é ainda mais elevado que para a métrica ISI. Em nossa implementação, foi necessária a paralelização da computação para análises extensivas, que analisam de maneira sistemática uma conjunto de cenários. Conforme mencionado na Seção 2.2.3.3, a métrica tem limitações para lidar com atrasos, como verificamos em nossas simulações, ao compararmos, por exemplo, as Figuras 3.5o e 3.6o.

Conforme o esperado, a complexidade das métricas de similaridade entre trens de *spike* aumenta à medida que mais "peso"é atribuído à ocorrência individual de um potencial de ação. Essa complexidade é, inevitavelmente, traduzida no aumento do custo computacional necessário ao cálculo da matriz de similaridade S. As métricas de similaridade baseadas na taxa de disparos são menos complexas que as com escala temporal variável, que, por sua vez, são menos complexas que as baseadas nas relações temporais entre *spikes*. Neste trabalho, avaliamos se esse aumento de complexidade e, consequentemente, a consideração de padrões de interação que levam em conta a estrutura temporal dos trens de *spike*, se traduzem em benefícios significativos na detecção de ANs ou se uma alternativa mais adequada seria a opção por metodologias mais simples, com eficácia similar.

A dependência paramétrica de algumas métricas dificulta também a comparação sistemática entre elas, principalmente se considerarmos que o tempo de processamento de grandes quantidades de dados já é elevado. Pudemos verificar também que a utilização de métricas

que levam em conta a relação temporal entre *spikes* limita nossas simulações a intervalos de tempo menores, já que elas requerem mais poder computacional. Assim, notamos que a análise de um conjunto muito grande de células por um longo intervalo de tempo sugere o uso de métricas de similaridade baseadas na taxa de disparo, já que o processamento necessário para a análise dos padrões mais complexos presentes na estrutura temporal dos trens de *spike* pode ser inviável.

Uma alternativa promissora seria o uso de arquiteturas paralelas de processamento (GPUs, *graphics processing units*, na literatura em inglês). Por exemplo, Rossant *et al.* (2010), Rossant *et al.* (2011) utilizou GPUs para determinar os parâmetros de modelos computacionais de redes neurais com base em dados eletrofisiológicos e obteve um sistema 80 vezes mais rápido; Cao *et al.* (2009) utilizou GPUs para mineração de dados de neurônios e obteve um ganho de velocidade de 15 vezes; Raimondo *et al.* (2012) relatam uma melhora de 25 vezes no processamento de sinais de eletroencefalograma; De Camargo *et al.* (2011) simularam redes com mais de 200 mil neurônios com uma velocidade 40 vezes maior que uma CPU padrão.

Entretanto, os resultados retratados nesta dissertação analisam apenas um valor de escala temporal para os dados simulados e ressaltamos que, variações nos parâmetros podem produzir diferenças significativas (KRUSKAL *et al.*, 2007). Devido a essa complexidade computacional, os resultados desta dissertação referem-se a um subconjunto de valores dos diversos parâmetros aceitos por cada algoritmo, e que foram escolhidos com base em suas propostas originais e relevância ao problema abordado. Apesar deste cuidado, ressaltamos que variações nesses parâmetros podem produzir diferenças significativas nas conclusões apresentadas. Mais especificamente neste trabalho, utilizamos *bins* de 10 ms, de maneira análoga ao utilizado por Pais-Vieira *et al.* (2013b), já que analisamos dados reais de animais envolvidos na mesma tarefa comportamental. Além disso, empregamos os parâmetros das métricas de similaridade com escalas temporais variáveis de maneira que $\sigma = \frac{1}{c_v} = \frac{Tamanho\ do\ bin}{\sqrt{12}}$, tomando por inspiração Kruskal *et al.* (2007) e Humphries (2011), que empregam uma relação parecida, mas, baseada os valores de ISI na população.

Cabe ressaltar também que, de maneira geral, é esperado que o desempenho das métricas de similaridade melhorem com trens de *spike* de maior duração, já que, mais informação está disponível (PAIVA *et al.*, 2010). Entretanto, na prática, para tarefas comportamentais, esse valor costuma ser inferior a um segundo (PAIVA *et al.*, 2010). Por isso, uma das motivações deste estudo foi o de analisar diferentes métricas de similaridade para destacar as mais promissoras em detectar interações neurais com uma menor informação disponível. Observamos também que a maioria das métricas não modela as interações inibitórias.

4.1.2 Métodos de classificação

As observações reunidas sobre os algoritmos de detecção de comunidades foram:

- PCA e SVD associados ao ICA: O PCA possui um melhor desempenho a medida que a interação entre os neurônios se torna mais forte e seu uso não é adequado para processos pontuais (SOLO; PASHA, 2012). Como a maioria das correlações das células são de baixa ordem de magnitude (LONG; CARMENA, 2011), uma opção é o uso de *bins* maiores ou trens de maior duração, o que dificulta seu uso para o tratamento de efeitos na atividade neuronal de escala temporal rápida. Outra possibilidade seria a incorporação de rotinas adaptadas para processos pontuais, como a de Solo e Pasha (2012). Além disso, conforme mencionado na Seção 2.3, tanto o PCA quanto o SVD requerem que a matriz de similaridade S seja simétrica, o que exclui análises de interações direcionadas.

- Modularidade: A Modularidade apresenta um limite de resolução (FORTUNATO; BARTHELEMY, 2007), ou seja, tem uma tendência a agrupar comunidades distintas em comunidades maiores. Além disso, a Equação 2.29, utilizada para o modelo nulo, tem como premissa que as redes são grandes e necessita de uma correção para redes menores, da ordem de poucas dezenas (NEWMAN, 2006). Outras equações para o modelo nulo podem ser testadas e impactar o desempenho, como um modelo que considera grafos direcionados (LEICHT *et al.*, 2007) e uma em que a equação captura a tendência de nós similares estarem agrupados (LIU *et al.*, 2012). Além disso, seu uso no processamento de dados eletrofisiológicos, apresentado por Humphries (2011), não é capaz de detectar adequadamente comunidades que compartilham neurônios, como pode ser verificado pela Figura 3.21, em que o desempenho atinge um patamar inferior em relação à Figura 3.19. Nesses casos, os neurônios em comum foram atribuídos apenas a algumas das ANs detectadas. Como trabalho futuro, sugerimos a incorporação de novos propostas que solucionam o problema de elementos em comum, como a apresentada por Nicosia *et al.* (2008).

  A Modularidade requer também a definição de uma matriz de similaridade S com diagonais nulas, uma vez que, por hipótese, uma célula não pode estar conectada a si mesma e nenhum elemento $s_{i,j}$ pode ser negativo, já que o método não é capaz de trabalhar com conexões negativas. Para finalizar, como o algoritmo se dispõe apenas a particionar um grafo, notamos que, de maneira geral, elementos independentes não foram isolados, ou seja, quando as comunidades foram detectadas, todos os elementos foram alocados a alguma das comunidades, como pode ser verificado, por exemplo, na Figura 3.17, em que o desempenho para a classificação dos neurônios saturou no patamar de 50%, uma vez que as 10 células independentes foram alocadas a alguma

das assembleias. Outro exemplo é a Figura 3.19, em que a saturação de desempenho fica no patamar de 80%.

- Agrupamento hierárquico: Neste trabalho utilizamos a mesma solução empregada por Kreuz *et al.* (2007), que consiste na implementação pelo software MATLAB®. A clusterização utilizada é do tipo exclusiva, e, portanto, assim como a Modularidade também possui limitações ao lidar com ANs não disjuntas, uma vez que realiza agrupamentos aos pares. Além disso, ele não lida com métricas de similaridade não simétricas, o que naturalmente exclui a análise de interações direcionadas (métrica proposta por Billeh *et al.* (2014)). Suas limitações mais conhecidas dizem respeito à baixa robustez a pequenas variações e a *outliers*. Algumas soluções a este problema foram propostas, como o trabalho de Balcan *et al.* (2014).

  Ademais, uma maneira geral de particionamento do dendrograma em diferentes comunidades permanece em discussão para aplicações do algoritmo em contextos variados (ESTIVILL-CASTRO, 2002) e propostas aplicadas a registros eletrofisiológicos extracelulares ainda são pouco exploradas. Notamos que a distinção entre a ocorrência de ANs e agrupamentos de células independentes nem sempre é bem delimitada para um observador sem conhecimento prévio da estrutura dos dados, como pode ser verificado, por exemplo, na Figura 3.23, o que se traduz num desempenho ruim para implementações automáticas.

  As propostas de particionamento automático utilizadas neste trabalho não se mostraram adequadas devido a alguns fatores como o dendrograma não exibir as relações pré-estabelecidas entre os elementos (por exemplo, o uso da métrica de similaridade de Hamming para a taxa média de disparos de 64 *spikes/segundo*, nas Figuras 3.28 e 3.30) ou como a baixa robustez a pequenas variações na definição do ponto de inflexão, definido pelo parâmetro ϵ (Figura 3.22 e Seção 2.3.4.4). Isso se deve ao fato de definirmos a ocorrência do ponto de inflexão nos índices de validação utilizando apenas dois pontos de um conjunto discreto (expressões descritas na Seção 2.3.4.4). Dessa maneira, pequenas alterações no valor de quaisquer destes pontos poderiam variar consideravelmente a localização do ponto de inflexão e, portanto, alterar o resultado. Não avaliamos os efeitos da variação sistemática deste parâmetro para cada uma das métricas de similaridade.

- Estabilidade de uma partição: O algoritmo também se baseia na modelagem de ANs segundo um modelo de grafos. Delvenne *et al.* (2010) apresenta a Estabilidade de uma partição como uma generalização da Modularidade, afirmando que o particionamento retornado pela Modularidade pode ser ora muito fino, ora muito grosseiro e em parti-

ções transitórias, que podem não corresponder à dinâmica do sistema. Dessa maneira, verificamos que esses algoritmos compartilham limitações. Notamos que a Estabilidade de uma partição também falha ao lidar com topologias em que as assembleias compartilham neurônios (Figura 2.1), como pôde ser verificado na Figura 3.38.

Assim como para a Modularidade, ele visa particionar um grafo e, portanto, uma vez que comunidades sejam detectadas, cada neurônio é alocado a alguma delas. Dessa forma, ele apresenta limitações ao lidar com células independentes, o que foi verificado, por exemplo, para os 3 pontos em que o desempenho da classificação dos neurônios atingiu 80% e a detecção de ANs 50%, na Figura 3.34. Nestes casos 3 assembleias foram detectadas, sendo 2 delas correspondentes às artificialmente inseridas, e a terceira erroneamente detectada e composta por todos os neurônios independentes. Uma possibilidade para contornar este problema seria a modificação do particionamento automático, neste trabalho, definida como à correspondente ao maior *plateau* em que VI=0 ao longo do tempo de Markov. Uma modificação nesta etapa poderia também diminuir os falsos positivos para a detecção de ANs nos cenários em que as células eram independentes entre si e as assembleias eram compostas por poucas células (Figura 3.34a, por exemplo). Ressaltamos também que o custo computacional desta metodologia é significativamente maior quando comparada às demais, e sua utilização para simulações de um grande conjunto de cenários pode ser inviável.

Conforme brevemente mencionado na Seção 2.3.7, a noção de comunidade não é um conceito bem definido. Estivill-Castro (2002) mostra que a noção pode ser construída de duas maneiras: de cima para baixo, *top-down* na literatura em inglês, em que o objetivo é segmentar uma população heterogênea em um agrupamentos mais homogêneos; e de baixo para cima, *bottom-up* em inglês, quando se deseja encontrar grupos relacionados entre si por algum "critério natural de similaridade".

O PCA e o SVD (seguidos ou não pelo ICA) podem ser atribuídos a uma metodologia do tipo *bottom-up*, pois buscam por interações mais significativas do que em caso de elementos independentes entre si. Já a Modularidade, o Agrupamento Hierárquico e a Estabilidade de uma partição são metodologias do tipo *top-down*, pois são estratégias de particionamento [1].

[1] Neste trabalho utilizamos as palavras *top-down* e *bottom-up* para se referirem à maneira que as comunidades emergem a partir de um conjunto de dados. No entanto, para o Agrupamento Hierárquico essas nomenclaturas são utilizadas em um contexto ligeiramente diferente, para categorizar as estratégias de agrupamento dos elementos. A estratégia *top-down* (ou de divisão) enxerga os elementos como membros de uma grande comunidade e estabelece relações em agrupamentos menores, indo recursivamente até o nível dos elementos. Já a *bottom-up* (ou de aglomeração) encara cada elemento como uma comunidade e, interativamente, agrupa os elementos mais similares entre si até que todos sejam reunidos em uma única comunidade (GUPTA, 2005). Neste trabalho, estabelecemos as relações entre os elementos por meio de uma estratégia *bottom-up*.

O critério de Marcenko e Pastur (1967) (Equação 2.24), utilizado na detecção de ANs para o PCA/SVD, faz uso de um modelo estatístico para contabilizar o número de comunidades no conjunto de dados. Já a Modularidade, o Agrupamento Hierárquico e a Estabilidade de uma partição retornavam o resultado somente após a otimização iterativa de um determinado índice. Para a Modularidade, foram utilizadas diversas execuções do algoritmo *K-means*, para o Agrupamento Hierárquico, os índices de Davies-Bouldin e Silhouette e, para a Estabilidade de uma partição, o índice VI.

Cabe ressaltar que o resultado da aplicação destes algoritmos deve ser julgado segundo um critério de validade que possua uma relação intrínseca com os dados a serem analisados e não seja apenas um índice para ser otimizado iterativamente. Isso é importante pois determina o quanto podemos confiar que uma comunidade detectada está de fato presente, evitando a ocorrência de falsos positivos. Índices otimizados de maneira iterativa se traduzem na formalização de um conceito abstrato que define o conceito de comunidade e, essa noção pode não ser aplicável em um determinado problema (ESTIVILL-CASTRO, 2002). Podemos visualizar, por exemplo, que os índices de Davies-Bouldin e Silhouette associados ao Agrupamento Hierárquico não se mostraram adequados para a detecção de ANs (como exemplificado na Figura 3.22).

Além disso, os modelos utilizados neste trabalho se baseiam em heurísticas que encontram soluções aproximadas ao analisar a interação de objetos aos pares. No entanto, a avaliação das interações entre os elementos de uma AN poderia ser feita levando em conta grupos maiores segundo outros critérios (ESTIVILL-CASTRO, 2002).

Finalmente, ressaltamos que a qualidade de um algoritmo de detecção de comunidades deve levar em conta não apenas sua acurácia, mas também o custo computacional associado e quanto um aumento em sua complexidade pode se traduzir em melhorias de desempenho.

4.2 Discussão dos resultados para conjuntos de dados simulados

Nesta seção discutimos os resultados apresentados na Seção 3.1, em que aplicamos diferentes métricas de similaridade e métodos de classificação em conjuntos artificiais de trens de *spike* que simulavam a presença de ANs. Neste estudo escolhemos analisar o impacto no desempenho segundo os seguintes fatores: variação da taxa média de disparos da população, número de neurônios que compunham cada AN, a disposição dos neurônios nas ANs (Figura 2.1), presença ou ausência de atrasos de fase e, finalmente, o modelo utilizado na geração dos dados. Nossa análise de desempenho é dividida entre capacidade de reconhecer uma assembleia, representada pela Equação 2.37 e capacidade de distinguir os neurônios que a compõe, na Equação 2.4.2.

4.2.1 Desempenho dos algoritmos com diferentes taxas médias de disparo

Uma dificuldade inicial encontrada no *spike sorting* de registros extracelulares é a detecção de unidades com poucos disparos ao longo do experimento. Espera-se que à medida que a taxa média de disparos diminua existam menos ocorrências de potenciais de ação, o que dificultaria a extração de informações para a detecção das comunidades. Dessa maneira, o objetivo desta etapa foi verificar a sensibilidade do conjunto de métodos de processamento explorados neste trabalho quanto à taxa média de disparos da população. Duas taxas médias de disparos foram utilizadas: a primeira, extraída de uma seção experimental de um rato *Long-Evans* engajado na tarefa comportamental descrita na Seção 2.5, de 22,37 *spikes/segundo*; a segunda, de um valor extrapolado a 64 *spikes/segundo*.

4.2.1.1 Discussão: Variação do desempenho do ICA com diferentes taxas médias de disparo

As comunidades são detectadas por meio da comparação dos autovalores da matriz de similaridade S, que devem ser maiores que λ_{max}, conforme descrito na Equação 2.24. Observamos que, para as simulações realizadas com as métricas de similaridade distintas da correlação, ainda que os trens gerados fossem independentes, um autovalor era sempre superior a este limiar, levando o algoritmo a, erroneamente, apontar a existência de uma AN em todos os cenários. Isso se deve ao fato de termos adaptado métricas que quantificam a distância entre dois trens de *spike* para métricas de similaridade, como descrito na Seção 2.2.4. Como pode ser verificado na distribuição de Marcenko e Pastur (1967) (Equação 2.24), a distribuição recebe um parâmetro σ, que é a variância dos elementos da matriz. Para a correlação, σ é calculado a partir da taxa média de disparos das células. Assim, à medida que passamos a utilizar outras métricas de similaridade, o mais adequado seria o uso de técnicas de *bootstraping* para o levantamento de σ e, portanto, de λ_{max}. No entanto, o custo computacional associado ao *bootstraping* de algumas das métricas de similaridade torna-se proibitivo para simulações sucessivas. Nossos resultados indicam que o uso de σ tal qual como o calculado para a correlação produz poucos falsos positivos em cenários com mais de uma AN e motiva, portanto, seu uso. Isso pode ser verificado e.g. para as métricas de Sicronização de Evento, ISI e distância SPIKE (Figuras 3.4m, 3.4n, e 3.4o, respectivamente).

Em nossas simulações, o aumento da taxa média de disparos implica um aumento das interações entre os neurônios de uma AN. Nossos resultados indicaram uma relação direta entre as interações e o desempenho do ICA. Notamos, por exemplo, ao compararmos a Figura 3.2 com a 3.3 que, para a correlação e para as métricas de similaridade baseadas em relações temporais (Sincronização de Evento, ISI e distância SPIKE), há um aumento na região em que registramos a detecção de duas ANs. Para o cosseno, houve um grande número de ANs detectadas espuriamente quando a taxa de disparos era baixa.

Esses resultados poderiam sugerir o uso da combinação entre ICA com as métricas de similaridade baseadas nas relações temporais entre *spikes* visando aprimorar a detecção do número de ANs em registros eletrofisiológicos extracelulares coletados experimentalmente. No entanto, observamos que estes cenários nem sempre estão associados a um elevado desempenho na classificação dos neurônios, como pode ser visualizado, por exemplo, nas Figuras 3.4p, 3.4q e 3.4r. Além disso, os melhores desempenhos foram observados para a taxa média de disparo aumentada, o que, na prática, não é muito promissor, uma vez que taxas de disparo dessa ordem de magnitude são pouco frequentes. Acrescentamos também que o tempo de processamento destas métricas de similaridade tornam sua utilização pouco viável à medida que o número de células registradas simultaneamente aumenta.

As métricas de similaridade de Victor e Purpura e Van Rossum também apresentaram melhoria de desempenho quando as ANs contém um número maior de neurônios (regiões de máxima inibição e excitação para a interação das assembleias, nas Figuras 3.5c e 3.5i) com o aumento da taxa média de disparos. No entanto, além de corresponderem a um cenário pouco usual com relação à taxa média de disparos, elas trazem uma restrição adicional, pois capturam apenas ANs compostas por um maior número de células.

Pudemos observar também que quando as variações na taxa média de disparos das ANs tinham magnitudes similares, os pesos das PCs eram mais parecidos. Esse fato, aliado ao limiar fixo que definia se um neurônio pertencia ou não a uma AN (Equação 2.25), aumentou o número de falsos positivos e falsos negativos, justificando a queda de desempenho observada nas diagonais das figuras. O desempenho poderia ser melhorado por meio da inspeção dos pesos de cada PC (w_k) por um experimentador.

Especificamente para a correlação, que consiste em uma métrica baseada na taxa média de disparos, uma situação com 64 *spikes/ segundo* e com *bins* de 10 ms pode também ser encarada como um conjunto de neurônios registrados por um intervalo de tempo maior, com *bins* de maior duração (Figuras 3.3 e 3.5). Isso evidencia a observação levantada na Seção 4.1.2, de que este algoritmo apresentaria melhores resultados para uma análise numa escala temporal maior. Além disso, nossos resultados confirmaram a afirmação de Santos *et al.* (2011), que uma variação mínima na taxa média de disparos é necessária para uma correta identificação do número de ANs. Pudemos complementá-la, incluindo também interações inibitórias.

4.2.1.2 Discussão: Variação do desempenho da Modularidade com diferentes taxas médias de disparo

Nossas simulações sugerem que a Modularidade não seria o algoritmo mais apropriado para a caracterização de ANs. Para o cenário em que a taxa de disparos simulada era de 22,37

spikes/segundo, numericamente igual à obtida experimentalmente, obtivemos um resultado não nulo apenas quando utilizamos o cosseno como métrica de similaridade. A diferença entre os pares de Figuras (3.16, 3.17) e (3.18, 3.19) insinuam que há uma mínima taxa de disparos para que o método pare de apontar falsos positivos nas regiões centrais das figuras, onde as populações simuladas são praticamente independentes entre si.

A evidência do efeito da taxa média de disparos é ainda maior para ANs maiores, notada ao compararmos as Figuras 3.17 e 3.19. Os falsos positivos desapareceram, mas, apesar do melhor desempenho na detecção de comunidades a uma taxa média de disparos maior, de 64 *spikes/segundo*, principalmente para populações maiores (Figura 3.19), ressaltamos que taxas médias de disparo desta magnitude não são compatíveis com os dados experimentais considerados neste estudo.

4.2.1.3 Discussão: Variação do desempenho do Agrupamento Hierárquico com diferentes taxas médias de disparo

Conforme mencionado na Seção 4.1.2, os particionamentos automáticos baseados nos índices de validação não apresentaram resultados robustos. No entanto, a observação dos dendrogramas montados para o Agrupamento Hierárquico permitiram analisar visualmente o comportamento das métricas de similaridade em diferentes formas de *Linkage* com a taxa média de disparos.

Nossas observações indicam que as interações entre os neurônios das ANs artificiais com uma taxa média de disparos de 22,37*spikes/segundo* são pouco claras e que existiram muitas associações espúrias com as células independentes nos dendrogramas exemplificadas pela correlação (Figura 3.23). Além disso, a diferenciação do caso em que todas as células são independentes não é evidente.

O aumento da taxa média de disparos permite que os dendrogramas revelem as estruturas das ANs para algumas das métricas de similaridade utilizadas. A métrica de Hamming não exibiu resultados promissores nem para casos em que a taxa média de disparos era baixa nem quando foi aumentada artificialmente (Figuras 3.28 e 3.30). No entanto, acreditamos que seu uso seja mais apropriado em trens com uma baixa taxa média de disparos e não homogêneos, sendo, portanto, capazes de identificar a ocorrência de disparos diretamente associados a um determinado evento.

Nas Figuras 3.28 e 3.30, as métricas de correlação, Victor e Purpura, Van Rossum, ISI e distância SPIKE permitiram observar de maneira clara a separação das comunidades. A divisão fica mais evidente se combinarmos as visualizações dos dendogramas referentes à ligação simples, que parecem indicar mais adequadamente possíveis grupos de células inde-

pendentes, com os referentes à ligação média simples, que parecem apontar de maneira mais clara os grupos presentes em um conjunto de dados. No entanto, ressaltamos que dendrogramas como os exibidos nas Figuras 3.28 e 3.30 são referentes a um cenário em que tanto a taxa média de disparos quanto as interações entre as células de uma AN assumem valores limites, a fim de evidenciarmos a separação dos conjuntos. Num caso real, com ANs com uma taxa média de disparos mais baixa e interações mais fracas, a diferenciação das comunidades nos dendrogramas é mais difícil.

Portanto, o dendrograma apresenta-se como uma ferramenta para exibir visualmente as interações entre as células. Foi possível perceber, por exemplo, que a sensibilidade das métricas de similaridade às interações inibitórias é maior do que em relação às excitatórias, já que, para a maioria das métricas, a separação visual dos ramos correspondentes às ANs com interações inibitórias foi mais evidente. Dessa maneira, nossos resultados sugerem que o uso do dendrograma é interessante para evidenciar os padrões presentes nos dados, mas que a caracterização das ANs com taxas de disparo obtidas experimentalmente não seria recomendada, dada a dificuldade em diferenciar a estrutura das ANs da observada para células independentes (Figuras 3.27 e 3.29).

4.2.1.4 Discussão: Variação do desempenho da Estabilidade de uma partição com diferentes taxas médias de disparo

A análise dos dados simulados nos possibilitou observar que a Estabilidade de uma partição, da maneira utilizada neste trabalho, não apresentou melhoria no desempenho com o aumento da taxa média de disparos para nenhuma das métricas de similaridade analisadas, à exceção da métrica de Billeh. Para esta métrica, nossas simulações mostraram que o aumento de 22,37 *spikes/segundo* para 64 *spikes/segundo* resultou numa melhoria de desempenho concentrada nas regiões em que as interações entre as células das ANs são de excitação. No entanto, as Figuras 3.33 e 3.34 apontam que este algoritmo não seria o mais adequado para o processamento de um conjunto de dados obtidos experimentalmente, já que foi possível observar um elevado número de falsos positivos para a detecção de ANs quando as células eram independentes entre si.

Além disso, apesar de a métrica de similaridade de Billeh alterar sua computação levando em consideração as interações de inibição entre as células que compõem uma AN, como descrito na Seção 2.2.2.4, não observamos uma melhoria de desempenho simultâneo para a detecção das ANs e para a classificação das células, mesmo com o aumento da taxa média de disparos (Figura 3.36).

4.2.2 Desempenho dos algoritmos com ANs de tamanhos diferentes

Nesta próxima etapa, analisamos o efeito no desempenho dos algoritmos à medida que mais neurônios participavam das ANs. De maneira geral, padrões retornados por comunidades maiores são mais facilmente detectáveis já que as diferenças em relação aos demais elementos se torna mais evidente. Além disso, há um limite de resolução intrínseco a todos os métodos, pois, inevitavelmente, precisamos distinguir quando as interações são significativas ou apenas acaso. Assim, o compromisso entre significância e sensibilidade estará sempre presente. Nesta seção, analisamos o desempenho de ANs independentes. No primeiro cenário, cada AN era composta por 10 células e, no segundo, por 25 células (Seção 3.1.1).

4.2.2.1 Discussão: Variação do desempenho do ICA com ANs de tamanhos diferentes

A comparação da Figura 3.2 com a Figuras 3.4 indicou que o desempenho do ICA melhora à medida que aumenta o número de neurônios que participam das ANs para as métricas de similaridade baseadas em relações temporais (Sincronização de Evento, ISI e distância SPIKE) e para a correlação. Esse fato pode ser justificado pois, como os resultados de identificação de ANs estatisticamente significativos estão associados com um limiar baseado nos autovalores (neste trabalho correspondendo ao limite superior da distribuição de Marčenko- Pastur na Equação 2.24), e o autovalor indica a quantidade de variância em uma dada direção, teremos, portanto, uma maior variância à medida que o número de elementos aumenta.

Assim, nossas simulaçõs sugerem que o uso do ICA associado a métricas de similaridade que capturam padrões temporais mais ricos (Sicronização de Evento, ISI e distância SPIKE) possui um viés para a detecção de ANs maiores. Por sua vez, para a correlação, existe uma relação entre o aumento da escala temporal em questão (correspondente a aumentarmos o tamanho dos *bins*) e a caracterização de comunidades menores: quanto maior a escala temporal, maior a resolução do algoritmo e vice-versa. Logo, concluímos que o experimentador, tendo ciência dessa relação de compromisso, pode melhorar o desempenho de suas análises ao utilizar o ICA caso esteja disposto a trabalhar com fenômenos de dinâmica mais lenta.

4.2.2.2 Discussão: Variação do desempenho da Modularidade com ANs de tamanhos diferentes

Foi possível observar que a Modularidade apresenta um desempenho significativamente melhor para ANs maiores, o que pode ser verificado ao compararmos as Figuras 3.17 e 3.19. Para ANs menores, houve um bom desempenho na detecção de comunidades nos casos em que ao menos uma das ANs interagia segundo a maior inibição possível. Para as maiores, o desempenho aumentou também em regiões em que as duas ANs possíam interações de

excitação. Isso evidencia a afirmação de Fortunato e Barthelemy (2007), já mencionada na Seção 4.1.2 sobre as limitações de resolução do algoritmo, que tende a agrupar as comunidades menores. Apesar de favorecer comunidades maiores, os resultados obtidos se mostraram confiáveis (sem falsos positivos) apenas num cenário em que a taxa de disparos foi aumentada artificialmente e que, portanto, não são adequados ao cenário experimental analisado neste trabalho.

4.2.2.3 Discussão: Variação do desempenho do Agrupamento Hierárquico com ANs de tamanhos diferentes

Como exibido na Figura 3.22, o particionamento automático avaliado neste trabalho não apresentou resultados satisfatórios. Nossas observações se basearam no formato dos dendrogramas obtidos para as métricas de similaridade analisadas. Particularmente para as topologias simuladas neste trabalho, com exceção do maior número de elementos que compõe as folhas, não foi possível observar diferenças significativas no formato geral dos dendrogramas à medida que aumentamos o tamanho das ANs.

Entretanto, relembramos que as diferenças entre os dendrogramas são mais pronunciadas ao compararmos as formas distintas de *Linkage* dependendo da estrutura dos dados, conforme mencionado brevemente na Seção 2.3.4. De maneira geral, os métodos clássicos de *Linkage* não possuem uma boa resposta na presença de comunidades de formato não esférico, de *outliers* ou ao lidar com cenários em que uma das comunidades é muito maior do que outras (GUHA *et al.*, 1998). Algumas propostas que endereçam alguns destes problemas são conhecidas, como revisado por Gupta (2005), e, o estudo da viabilidade da aplicação destas propostas a dados eletrofisiológicos extracelulares é sugerida como trabalho futuro.

4.2.2.4 Discussão: Variação do desempenho da Estabilidade de uma partição com ANs de tamanhos diferentes

Conforme a observação de Delvenne *et al.* (2010), o algoritmo é uma generalização da Modularidade ao analisarmos o tempo de Markov em $t = 1$. Assim, realizamos uma comparação entre o desempenho da Modularidade com o da Estabilidade de uma partição associados a diferentes métricas de similaridade, para verificar se a dinâmica das partições possui variações com o aumento do tempo de Markov. A Modularidade apresentou um elevado número de comunidades espúrias detectadas quando as células eram independentes entre si na taxa média de disparos de 22,37 *spikes/segundo* (Figuras 3.16 e 3.18). Já na Estabilidade de uma partição, a ocorrência de falsos positivos estava associada ao tamanho das ANs (Figuras 3.33 e 3.34), e diminuiu quando as ANs aumentaram (Figura 3.35).

Assim, nossos resultados sugerem que o aumento de complexidade e de custo computacional trazido pela Estabilidade de uma partição melhora os resultados à medida que aumenta o número de neurônios que compõem uma AN. Nossas simulações mostraram uma melhoria de desempenho quando as duas ANs possuíam interações excitatórias entre seus elementos além de uma queda considerável no número de falsos positivos para a taxa média de disparos obtida experimentalmente, de 22,37 *spikes/segundo* (Figura 3.35). Dessa maneira, as simulações apontam que a utilização deste método ao utilizarmos o particionamento automático no limiar $VI = 0$ (Seção 2.3.5) também apresenta um limite de resolução. Mais estudos são necessários a fim de determinar um limiar a partir do qual se tenha confiança que ANs detectadas são reais e não falsos positivos.

4.2.3 Desempenho dos algoritmos em ANs considerando diferentes disposições dos neurônios

A seguir, nosso objetivo foi caracterizar os algoritmos quanto a sua capacidade de detectar as ANs à medida que elas passavam a compartilhar neurônios. Isso é relevante pois existem evidências sugerindo que alguns neurônios podem participar simultaneamente de várias ANs. As topologias analisadas neste trabalho foram representadas na Figura 2.1.

O resultado para a aplicação do ICA para as simulações utilizando o modelo de Poisson indicam que o desempenho do algoritmo para a detecção das ANs depende exclusivamente da variância em uma determinada e região e não está relacionada ao fato das assembleias possuírem ou não neurônios em comum. A detecção das ANs depende da variância ser grande o suficiente para ultrapassar o valor definido pela Equação 2.24. Por exemplo, caso muitos neurônios sejam compartilhados, mesmo que a taxa média de disparos seja aumentada, a variância resultante não será suficiente para que a assembleia seja detectada, como pode ser verificado no segundo e no quarto quadrante da Figura 3.9a. Por outro lado, mesmo quando as ANs não possuam neurônios disjuntos, o fato de elas serem compostas por um maior número de neurônios e interajam por interações mais fortes pode permitir sua detecção, como ocorre para o primeiro e para o quarto quadrantes da Figura 3.14a. O uso de métricas de similaridade baseadas nas relações temporais entre *spikes*, particularmente o ISI, permitiu a detecção de ANs em situações de compartilhamento de neurônios, com melhor desempenho em regiões em que as interações eram de inibição (por exemplo, Figura 3.8b). No entanto, deve ser ressaltado que a identificação de uma AN de maneira espúria permanece e que o desempenho para a classificação automática dos neurônios não foi satisfatório.

Nossas análises mostraram também, como mencionado na Seção 4.1.2, as limitações dos demais algoritmos (Modularidade, Agrupamento Hierárquico e Estabilidade de uma par-

tição) para lidarem com elementos em comum entre as ANs pré-estabelecidas, além de não identificarem de maneira robusta os neurônios independentes. No entanto, em nossas simulações foi possível observar que as ANs continuam, em muitos casos, a serem detectadas (Figuras 3.19a e 3.38a, por exemplo), ao custo de uma queda no desempenho para a classificação dos neurônios (Figuras 3.19d e 3.38b). Isso ressalta que um baixo desempenho na identificação do número de ANs não necessariamente implica um resultado similar para a classificação dos neurônios. Essa observação é importante porque salienta que, dependendo da aplicação, uma penalização nessa métrica pode ser aceitável, especialmente com o desenvolvimento de novas técnicas que estão aumentando o número de neurônios gravados simultaneamente. No entanto, como neste trabalho o interesse era verificar a maior ocorrência de ANs em instantes próximos à discriminação tátil e à recompensa em comparação a outros instantes, a classificação dos neurônios possuía muita relevância, uma vez que necessitávamos saber quais células participavam das assembleias.

4.2.4 Desempenho dos algoritmos em interações com atraso de fase

Observamos que o desempenho dos algoritmos em interações com atrasos de fase está diretamente relacionado ao efeito em cada uma das métricas de similaridade. O valor calculado pelas métricas de similaridade baseadas em taxas médias de disparos não é robusto a esse tipo de efeito. Isso pode ser observado, por exemplo, na associação da correlação com o ICA, nas Figuras 3.6a e 3.6d em comparação às Figuras 3.5a e 3.5d, já que as 5 células atrasadas, em alguns casos, foram agrupadas em uma nova AN e, em outros, classificadas como células independentes. A métrica de Hamming também não apresentou uma boa resposta ao atraso, como pode ser verificado na Figura 3.20d quando comparada à Figura 3.19d, pois observamos uma queda no desempenho para a classificação dos neurônios na região em que a AN2 possuía interações de inibição.

Já as métricas de Victor e Purpura, Cosseno e Van Rossum possuem parâmetros e seu resultado depende dos valores utilizados. Para a escala temporal avaliada neste trabalho, pudemos observar que o cosseno não apresentou robustez, como pode ser verificado pelas Figuras 3.20f e 3.20e quando comparadas às Figuras 3.19f e 3.19e, respectivamente. As métricas de Victor e Purpura e Van Rossum apresentaram uma ligeira queda de desempenho, como pode ser verificado ao compararmos os cantos do segundo e do quarto quadrante das Figuras 3.6c e 3.6i às Figuras 3.5c e 3.5i.

Salientamos a dificuldade de realizar análises sistemáticas já que estas métricas de similaridade são paramétricas e, variações nestes parâmetros podem acarretar variações nos resultados. Apesar dos parâmetros fornecerem flexibilidade ao experimentador e de ele poder modificá-los para abranger os atrasos, a identificação dos desvios de fase se torna difícil à

medida que aumentam as populações neuronais registradas, pois também aumenta a quantidade de dados. Muitas vezes a escolha do parâmetro é feita de maneira empírica, o que não necessariamente corresponde ao melhor resultado possível. É necessário atentar também ao fato de que a escolha de uma escala temporal pode limitar a observação de interações sinápticas mais complexas e em maior escala temporal, como as estabelecidas pelos potenciais sinápticos de longa duração ou pelos potenciais de depressão longa, *Long-Term Synaptic Potenciation* e *Long-Term Synaptic Depression* na literatura em inglês, respectivamente.

Neste contexto, implementações não paramétricas são favorecidas, como as métricas de similaridade baseadas nas relações temporais entre os *spikes*. Apesar de serem mais custosas computacionalmente, nossas simulações indicaram uma maior robustez ao efeito de atrasos de fase, quando comparamos as Figuras 3.6m, 3.6n e 3.6o com as Figuras 3.5m, 3.5n e 3.5o, respectivamente. A distância SPIKE foi a única a apresentar uma diferença quando os atrasos foram inseridos. Para o ISI e a Sincronização do evento, não observamos diferenças significativas entre as figuras.

Estudos demonstram que diferentes propriedades das respostas neurais contém informação que pode ser utilizada para a decodificação de parâmetros comportamentais (JACOBS *et al.*, 2009; HOUGHTON; VICTOR, 2010). Dessa maneira, dependendo do cenário experimental, a taxa média de disparos, ao invés do instante de ocorrência de cada *spike*, pode ser suficiente para caracterizar a formação de ANs, mesmo que alguns neurônios pertencentes às ANs não sejam detectados devido a um atraso temporal, por exemplo. Em aplicações em tempo real como e.g. interfaces cérebro-máquina, a utilização de métodos que empregam a taxa média de disparos pode ser preferível, sobretudo por normalmente resultarem em uma abordagem menos custosa computacionalmente.

4.2.5 Desempenho dos algoritmos segundo um modelo de disparos mais realista

A geração de trens de *spike* utilizando para sua descrição não somente a taxa média de disparos, mas também sua correlação cruzada apresentou limitações. A proposta de Macke *et al.* (2009) permite ao experimentador produzir conjuntos de dados artificiais de duas maneiras: na primeira, o usuário gera trens binários em que 1 corresponde a uma amostra em que foi registrada a ocorrência de um potencial de ação e 0, quando não houve uma ocorrência. Na segunda, o conjunto gerado já é dividido em janelas e os trens já representam a taxa média de disparos. Assim, para que fosse possível a produção de trens binários, extraímos os valores de uma população a 1 ms.

Porém, nesta escala temporal, as correlações entre as células obtidas foram muito baixas e o algoritmo não conseguiu gerar trens binários com esse grau de precisão na esta-

tística de segunda ordem. Portanto, precisamos aumentar a janela de análise para 10 ms, a mesma utilizada por Pais-Vieira *et al.* (2013b). Assim, pudemos analisar apenas as métricas de similaridade baseadas na taxa de disparo (Seção 2.2.1). No entanto, para *bins* de 10 ms os valores em módulo das correlações continuaram muito baixos, especialmente para correlações negativas. Não foi possível caracterizar com sucesso as interações inibitórias.

Como mencionado na Seção 2.1.2, o modelamento de trens de *spike* segundo um processo de Poisson é amplamente utilizado, mas essa simplificação não condiz com toda a variabilidade presente na atividade neuronal e, portanto, os resultados obtidos neste trabalho devem ser utilizados com cuidado ao se estudar modelos de processos pontuais mais complexos (PAIVA *et al.*, 2010). Nós buscamos complementar a análise apresentando a resposta dos algoritmos a trens artificiais com uma estrutura temporal mais rica, gerados pelo modelo proposto por Macke *et al.* (2009). Nossos resultados sugerem que o uso dessa modelagem é mais adequado para simulações numa escala temporal maior, em que a taxa média de disparos seja mais elevada, possibilitando o cálculo de correlações mais fortes entre os neurônios (DEOLINDO *et al.*, 2014). Foi possível observar pelas Figuras 3.7a, 3.10a e 3.15a que mesmo em cenários em que as interações entre os neurônios de uma AN era fraca, houve para o ICA um aumento no desempenho para a detecção de assembleias. Tais observações salientam que a negligência de interações mais complexas pode induzir falsas conclusões quando lidando com dados reais e reforça a necessidade de métodos mais realistas para testar as metodologias de processamento de sinais. Em trabalhos futuros, nós propomos análises semelhantes às apresentadas neste trabalho, porém, empregando outros modelos de trens de *spike*, como por exemplo o modelo de Poisson com período refratário e as propostas de Krumin e Shoham (2009) e Oweiss (2010).

4.2.6 Sumário dos Resultados

Na Tabela 4.1, a seguir, resumimos os principais achados das simulações realizadas neste trabalho. Em azul, demarcamos as associações entre métricas de similaridade e métodos de classificação que retornam um resultado nulo nos cenários analisados. Em laranja, marcamos a incompatibilidade entre a métrica de similaridade de Billeh e alguns algoritmos. Em verde, demarcamos as combinações mais promissoras, utilizadas na análise de registros eletrofisiológicos extracelulares. Buscamos também inserir marcadores (*) que possibilitassem uma visualização geral do aumento do custo computacional de cada uma das propostas. Demarcamos também as limitações devido ao parâmetro σ da Equação 2.24 para o ICA, e para lidar com células independentes e que pertencessem a mais de uma comunidade para a Modularidade, para o Agrupamento Hierárquico e para a Estabilidade de Markov.

Tabela 4.1 – Sumário das simulações realizadas neste trabalho ao combinarmos diferentes métricas de similaridade e métodos de classificação na caracterização de ANs.

Algoritmo / Métrica de Similaridade	**ICA[1]**	**Modularidade[2]**	**Clusterização Hierárquica[2]**	**Estabilidade de Markov[2**]**
Correlação	Maior variância: interações mais fortes e comunidades maiores. Melhor desempenho para maior escala temporal	Resultado nulo	Particionamento automático não apresentou desempenho favorável.	Resultado nulo
Hamming	Uma comunidade espúria.	Inibição mútua. Comunidades maiores. Mais adequado a taxas médias de disparo maiores.		
Victor e Purpura*	Excitação mútua ou inibição mútua associada a variações maiores na taxa média de disparos. Uma comunidade espúria.	Resultado nulo		
Convolução + Cosseno	Uma comunidade espúria.	Inibição mútua. Comunidades maiores. Taxas médias de disparo maiores. Resultados espúrios em interações fracas.		
Van Rossum	Mais adequado a excitação mútua ou inibição mútua associado a variações maiores na taxa média de disparos. Uma comunidade espúria.	Resultado nulo		
Sincronização de Evento**	Inibição mútua. Comunidades maiores. Mais adequado a taxas médias de disparo maiores. Uma comunidade espúria.			
ISI*	Inibição mútua. Comunidades maiores. Mais adequado a taxas médias de disparo maiores. Uma comunidade espúria.			
SPIKE****	Inibição mútua. Comunidades maiores. Mais adequado a taxas médias de disparo maiores. Uma comunidade espúria.			
Billeh*	-	-	-	Excitação mútua. Comunidades maiores. Resultados espúrios para baixas taxas médias de disparo.

1- Adaptação de parâmetro estatístico: Impacto no número de assembleias detectadas
2- Não lida com assembleias que compartilham neurônios. Não identifica neurônios independentes.
* Quanto maior o número de símbolos, maior o custo computacional.

4.3 Discussão dos resultados referentes à aplicação dos algoritmos a registros eletrofisiológicos extracelulares

A descrição dos registros eletrofisiológicos de animais engajados em uma tarefa de discriminação tátil mostrou que a aplicação do ICA associado à Correlação e ao ISI exibiram resultados consistentes entre si (Tabelas 3.2 e 3.3). O uso do ICA justifica-se já que foi o único método dentre os analisados a apresentar um bom desempenho nas simulações além de identificar com sucesso as células independentes. Nossas simulações permitiram constatar (Tabela 4.1) que o ISI tem um custo computacional maior que a correlação uma vez que calcula a similaridade entre trens de *spike* tomando por base a relação temporal entre os disparos. Nossos resultados sugerem que mesmo métodos mais simples podem caracterizar de maneira satisfatória a atividade das ANs, ou seja, confirmamos com essa análise que a utilização de um método menos custoso computacionalmente (ICA+Correlação) pode substituir, sem prejuízo, métodos mais elaborados. Este fato é relevante principalmente para análises em tempo real e com múltiplos animais (SHOKUR *et al.*, 2013; IFFT *et al.*, 2013; PAIS-VIEIRA *et al.*, 2013a; PAIS-VIEIRA *et al.*, 2015; RAMAKRISHNAN *et al.*, 2015). Pudemos verificar que ainda existe uma certa dificuldade em diferenciarmos as células às assembleias das células independentes.

Observamos também a existência de neurônios pertencentes à mesma região cortical engajados em ANs distintas, além de assembleias formadas por células de áreas diferentes. Notamos que as interações entre regiões corticais não se conservaram entre todos os animais, o que pode ser oriundo do desempenho diferente de cada animal na execução da tarefa, da velocidade distinta que cada um se movia em direção às barras, das pequenas diferenças anatômicas e das variadas estratégias que o cérebro de cada animal pode adotar para processar as informações recebidas.

5 Conclusões

Neste trabalho, realizamos uma análise comparativa do desempenho de diferentes formas de processamento de trens de *spike* na caracterização das ANs. Avaliamos propostas distintas para quantificar as interações entre as células (Seção 2.2) e para métodos de classificação (Seção 2.3), levantando também as limitações dessas etapas.

A implementação computacional foi modularmente estruturada em quatro etapas. A primeira, referente à obtenção dos dados, possibilitava a escolha entre a geração de trens de *spike* artificiais (Seção 2.1) ou registros experimentais. A segunda consistia na computação das interações entre as células por meio de uma das seguintes métricas de similaridade: Correlação, Hamming, Victor e Purpura, Cosseno, Van Rossum, Billeh, Sincronização de evento, ISI e distância SPIKE (Seção 2.2). A terceira era composta pela aplicação de um dos métodos de classificação: PCA associado ao ICA, Modularidade, Agrupamento Hierárquico e Estabilidade de uma partição (Seção 2.3). Finalmente, a quarta computava índices de desempenho (Seção 2.4) no caso de dados simulados ou exibia as ANs obtidas, no caso do processamento de dados reais. Cada uma das etapas foi associada a um arquivo de parâmetros, de maneira a flexibilizar as alterações.

No Capítulo 3, exploramos as possíveis combinações e diversas propriedades dos algoritmos selecionados em dados simulados e reais. As análises foram feitas tanto nos contextos originalmente propostos como em variações inéditas, o que permitiu destacar os pontos mais promissores de cada abordagem e desafios futuros.

As principais contribuições deste trabalho foram:

- Levantamento das hipóteses implícitas assumidas por cada uma das métricas de similaridade ou dos métodos de classificação no problema de detecção de ANs. A maioria das modelagens utilizadas neste trabalho foram originalmente propostas a outras classes de problemas e a adaptação ao contexto de processamento de registros eletrofisiológicos extracelulares pode não ser tão imediata. Métodos propostos em um contexto assumem dados segundo uma determinada estrutura e terão um desempenho ruim quando aplicados a um conjunto de estrutura diferente. Verificamos, por exemplo, que a Modularidade e a Estabilidade de uma partição, originalmente propostas para o particionamento de conjuntos de dados, apresentam limitações ao lidarem com células independentes.

- Análise do conceito de comunidade sobre o ponto de vista de diversos algoritmos. O conceito não é bem definido e, portanto, discutimos os pontos fortes e fracos de cada

um deles para nosso problema específico. A delimitação de um problema específico é importante pois a comparação dos algoritmos só é possível quando podemos definir se um determinado resultado é válido ou não segundo um critério que considera a estrutura dos dados.

- Ponderação dos resultados apresentados pelas diferentes métricas de similaridade e dos métodos de classificação segundo sua complexidade. Para as métricas de similaridade, por exemplo, algumas propostas utilizavam em sua descrição o tempo de ocorrência de cada potencial de ação, enquanto outras, se baseavam em escalas temporais maiores, e consideravam a taxa média de disparos. Também a métrica de Billeh modelava interações direcionais. Já para os métodos de classificação, por exemplo, a Estabilidade de uma partição era uma generalização da Modularidade. Ressaltamos ao longo deste trabalho que o aumento de complexidade de uma metodologia reflete num maior custo computacional de processamento e que esta variável é tão importante quanto o desempenho final na caracterização das ANs. Acreditamos que seria adequado ponderar a utilização de modelos mais complexos no processamento de dados experimentais levando-se em conta o resultado de simulações que mensurem a melhoria de desempenho que eles podem trazer. Assim, a complexidade do modelo seria aumentada apenas quando o impacto no resultado for significativo e não como regra geral. Essas considerações se tornam particularmente relevantes para aplicações de processamento em tempo real.

- Aplicação dos métodos mais promissores em um conjunto de registros eletrofisiológicos extracelulares de ratos engajados em uma tarefa de discriminação tátil, que envolve diversas áreas corticais no processamento da informação. Verificamos a formação de assembleias entre os neurônios dos córtices visual primário, somatossensorial primário, pré frontal e parietal posterior para todos os seis ratos analisados. Observamos também que as associações entre as áreas podem variar entre os animais, o que sugere que a circuitaria neuronal pode adotar diferentes estratégias de processamento da informação. Os resultados obtidos pelas duas metodologias foram similares e, portanto, recomendamos o uso da alternativa com menor custo computacional.

5.1 Perspectivas e propostas de extensões

Ressaltamos que as métricas de similaridade utilizadas neste trabalho mensuravam a interação entre 2 células de cada vez. Esta abordagem possui limitações, tendo em vista que o número de combinações aumenta exponencialmente e o número de células registradas tem aumentado com o desenvolvimento de novas técnicas de confecção de microeletrodos.

Abordagens alternativas que mensurem a atividade grupal de um maior número de células podem ser mais promissoras, como, por exemplo, a generalização multivariável da distância SPIKE (KREUZ *et al.*, 2011). Além disso, existem questionamentos sobre a abrangência do mapeamento de uma resposta cognitiva num espaço regido por relações de distância, já que existem fenômenos cognitivos que não são facilmente capturados por esta descrição matemática (HOUGHTON; VICTOR, 2010).

Frisamos também o desafio envolvido na realização de comparações sistemáticas entre os métodos, tendo em vista que eles apresentam uma grande quantidade de parâmetros. Optamos por variar alguns deles, e ressaltamos que os resultados podem sofrer alterações com modificações nos valores escolhidos. Sugerimos como trabalho futuro uma análise do impacto no resultado ao modificarmos cada um dos parâmetros. Nossos resultados estão de acordo com a afirmação de Chicharro *et al.* (2011), a de que o desempenho é mais sensível à escolha do método de classificação do que pela escolha da métrica de similaridade ou da forma do *kernel*.

Evidenciamos também as relações de compromisso entre a resolução espacial das ANs detectadas, que diminui à medida que diminui a escala temporal de interesse. Nossas simulações permitiram observar que os métodos ainda tem dificuldade de trabalhar simultaneamente com eventos de escala temporal rápida associados a comunidades pequenas com células que disparam pouco e salientam a importância do desenvolvimento de metodologias que enderecem este problema.

Nossas conclusões se baseiam em simulações que utilizam majoritariamente os dados simulados segundo um modelo de Poisson. Uma etapa fundamental do problema de detecção de comunidades é a compreensão da estrutura dos dados, pois fundamenta a interpretação dos resultados (ESTIVILL-CASTRO, 2002). Assim, sugerimos como trabalho futuro a utilização de modelos mais complexos para a construção dos trens de *spike*, principalmente aqueles capazes de trabalhar com processos pontuais(KRUMIN; SHOHAM, 2009). Além disso, nossas simulações consideravam que todas as células que pertenciam a uma AN interagiam da mesma maneira, ou seja, todas eram excitatórias ou inibitórias com interações de mesma magnitude. Sugerimos uma análise de assembleias em que as interações entre seus elementos seja variada.

5.2 Consideração final

Existem numerosas evidências que apontam que o código neural é distribuído (NICOLELIS *et al.*, 1997; HARRIS, 2005; BUZSáKI, 2010). Dessa maneira, o estudo de ANs é relevante já que auxilia o entendimento da organização conjunta dos neurônios no processamento da informação. A caracterização das assembleias apresenta uma série de desafios,

desde a aquisição (POLIKOV *et al.*, 2005; SCHWARZ *et al.*, 2014) até o processamento dos sinais (LEWICKI, 1998; PANZERI *et al.*, 2010; MULANSKY *et al.*, 2015). O presente estudo contribuiu na análise das propriedades e limitações de algumas metodologias utilizadas para a caracterização de ANs.

Referências

AGGARWAL, V.; MOLLAZADEH, M.; DAVIDSON, A. G.; SCHIEBER, M. H.; THAKOR, N. V. State-based decoding of hand and finger kinematics using neuronal ensemble and LFP activity during dexterous reach-to-grasp movements. *Journal of neurophysiology*, v. 109, n. 12, p. 3067–81, jun. 2013. ISSN 1522-1598. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3680811&tool=pmcentrez&rendertype=abstract>. Citado na página 3.

AMARASINGHAM, A.; CHEN, T.-L.; GEMAN, S.; HARRISON, M. T.; SHEINBERG, D. L. Spike count reliability and the Poisson hypothesis. *The Journal of neuroscience*, v. 26, n. 3, p. 801–9, jan. 2006. ISSN 1529-2401. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/16421300>. Citado na página 15.

BAKSHI, B. R. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE Journal*, v. 44, n. 7, p. 1596–1610, jul. 1998. ISSN 00011541. Disponível em: <http://doi.wiley.com/10.1002/aic.690440712>. Citado na página 31.

BALCAN, M. M.-f.; LIANG, Y.; GUPTA, P. Robust hierarchical clustering. *arXiv preprint arXiv:1401.0247*, v. 15, p. 35, 2014. Disponível em: <http://arxiv.org/abs/1401.0247>. Citado na página 115.

BEHRMANN, M.; GENG, J. J.; SHOMSTEIN, S. *Parietal cortex and attention*. 2004. 212–217 p. Citado na página 54.

BENTIVOGLIO, M. *Life and Discoveries of Camillo Golgi*. 1998. Disponível em: <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1906/golgi-article.html>. Citado na página 2.

BENTIVOGLIO, M. *Life and Discoveries of Santiago Ramón y Cajal*. 1998. Disponível em: <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1906/cajal-article.html>. Citado na página 2.

BILLEH, Y. N.; SCHAUB, M. T.; ANASTASSIOU, C. a.; BARAHONA, M.; KOCH, C. Revealing cell assemblies at multiple levels of granularity. *Journal of neuroscience methods*, Elsevier B.V., v. 236, p. 92–106, ago. 2014. ISSN 1872-678X. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/25169050>. Citado 10 vezes nas páginas 9, 10, 17, 22, 23, 39, 47, 49, 111 e 115.

BINGHAM, E.; HYVäRINEN, a. A fast fixed-point algorithm for independent component analysis of complex-valued signals. *Journal of Neural Systems,*, v. 10, p. 1–8, 2000. Citado na página 33.

BRUNEL, N.; ROSSUM, M. C. W. V. Lapicque's 1907 paper: From frogs to integrate-and-fire. *Biological Cybernetics*, v. 97, n. 5-6, p. 337–339, 2007. ISSN 03401200. Citado na página 13.

BUEHLMANN, A.; DECO, G. The neuronal basis of attention: rate versus synchronization modulation. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, v. 28, n. 30, p. 7679–86, jul. 2008. ISSN 1529-2401. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/18650344>. Citado na página 3.

BUZSáKI, G. *Rhythms of the Brain*. Oxford University Press, 2006. ISBN 9780195301069. Disponível em: <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195301069.001.0001/acprof-9780195301069>. Citado na página 3.

BUZSáKI, G. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, v. 68, n. 3, p. 362–85, nov. 2010. ISSN 1097-4199. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3005627&tool=pmcentrez&rendertype=abstract>. Citado 5 vezes nas páginas 3, 4, 16, 112 e 133.

CAO, Y.; PATNAIK, D.; PONCE, S.; ARCHULETA, J.; BUTLER, P.; FENG, W.-c.; RAMAKRISHNAN, N. Towards Chip-on-Chip Neuroscience: Fast Mining of Frequent Episodes Using Graphics Processors. p. 1–20, 2009. Disponível em: <http://arxiv.org/abs/0905.2200>. Citado na página 113.

CARMENA, J. M.; LEBEDEV, M. a.; CRIST, R. E.; O'DOHERTY, J. E.; SANTUCCI, D. M.; DIMITROV, D. F.; PATIL, P. G.; HENRIQUEZ, C. S.; NICOLELIS, M. a. L. Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS biology*, v. 1, n. 2, p. E42, nov. 2003. ISSN 1545-7885. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=261882&tool=pmcentrez&rendertype=abstract>. Citado na página 3.

CHAPIN, J. K.; MOXON, K. A.; MARKOWITZ, R. S.; NICOLELIS, M. A. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature neuroscience*, v. 2, n. 7, p. 664–70, jul. 1999. ISSN 1097-6256. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/10404201>. Citado na página 3.

CHICHARRO, D.; KREUZ, T.; ANDRZEJAK, R. G. What can spike train distances tell us about the neural code? *Journal of Neuroscience Methods*, Elsevier B.V., v. 199, n. 1, p. 146–165, 2011. ISSN 01650270. Disponível em: <http://dx.doi.org/10.1016/j.jneumeth.2011.05.002>. Citado na página 133.

CISEK, P.; KALASKA, J. F. Neural mechanisms for interacting with a world full of action choices. *Annual review of neuroscience*, v. 33, n. March, p. 269–98, jan. 2010. ISSN 1545-4126. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/20345247>. Citado na página 3.

COFER, D. *Neuron Basics*. 2002. Disponível em: <http://www.mindcreators.com/neuronbasics.htm>. Citado na página 5.

COLLINGER, J. L.; WODLINGER, B.; DOWNEY, J. E.; WANG, W.; TYLER-KABARA, E. C.; WEBER, D. J.; MCMORLAND, A. J. C.; VELLISTE, M.; BONINGER, M. L.; SCHWARTZ, A. B. High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet*, v. 381, n. 9866, p. 557–64, fev. 2013. ISSN 1474-547X. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3641862&tool=pmcentrez&rendertype=abstract>. Citado na página 3.

DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, v. 1, n. 2, p. 224–227, 1979. ISSN 0162-8828. Citado na página 45.

DAYAN, P.; ABBOTT, L. *Theoretical neuroscience*. [s.n.], 2001. Disponível em: <http://f3.tiera.ru/2/Cs_Computerscience/CsGn_Genetic,neural/DayanP.,AbbottL.F. Theoreticalneuroscience(2002)(432s).pdf.gz>. Citado 2 vezes nas páginas 15 e 21.

De Camargo, R. Y.; ROZANTE, L.; SONG, S. W. A multi-GPU algorithm for large-scale neuronal networks. *Concurrency Computation Practice and Experience*, v. 23, p. 556–572, 2011. ISSN 15320626. Citado na página 113.

DELVENNE, J.-C.; YALIRAKI, S. N.; BARAHONA, M. Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences of the United States of America*, v. 107, n. 29, p. 12755–12760, 2010. ISSN 0027-8424. Citado 4 vezes nas páginas 47, 49, 115 e 123.

DEOLINDO, C. S.; KUNICKI, A. C. B.; BRASIL, F. L.; MOIOLI, R. C. Limitations of principal component analysis as a method to detect neuronal assemblies. In: *2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)*. [s.n.], 2014. p. 24–30. ISBN 978-1-4799-6644-8. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7001808>. Citado 4 vezes nas páginas 10, 16, 110 e 127.

DRONGELEN, W. van. *Signal Processing for Neuroscientists*. Elsevier, 2006. ISBN 9780123708670. Disponível em: <http://www.just.edu.jo/FacultiesandDepartments/FacultyofEngineering/Departments/BiomedicalEngineering/Documents/NeuroscienceSyllabus.pdf>. Citado na página 18.

ENGEL, A. K.; MAYE, A.; KURTHEN, M.; KöNIG, P. Where's the action? The pragmatic turn in cognitive science. *Trends in cognitive sciences*, v. 17, n. 5, p. 202–9, maio 2013. ISSN 1879-307X. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/23608361>. Citado na página 3.

ESTIVILL-CASTRO, V. Why so many clustering algorithms — A Position Paper. *SIGKDD Explorations*, v. 4, p. 65–75, 2002. Citado 6 vezes nas páginas 29, 50, 115, 116, 117 e 133.

FALLAHATI, D. M.; BACKONJA, M.; EGHBALNIA, H.; ASSADI, A. H. Dynamic PCA for network feature extraction in multi-electrode recording of neurophysiological data in cortical substrate of pain. *Neurocomputing*, v. 44-46, p. 401–405, jun. 2002. ISSN 09252312. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0925231202003892>. Citado na página 31.

FELDT, S.; WADDELL, J.; HETRICK, V. L.; BERKE, J. D.; ZOCHOWSKI, M. Functional clustering algorithm for the analysis of dynamic network data. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, v. 79, p. 30–32, 2009. ISSN 15393755. Citado na página 9.

FELLOUS, J.-M.; TIESINGA, P. H. E.; THOMAS, P. J.; SEJNOWSKI, T. J. Discovering spike patterns in neuronal responses. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, v. 24, n. 12, p. 2989–3001, mar. 2004. ISSN 1529-2401. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2928855&tool=pmcentrez&rendertype=abstract>. Citado 2 vezes nas páginas 9 e 111.

FORTUNATO, S.; BARTHELEMY, M. Resolution limit in community detection. *PNAS*, v. 104, n. 1, 2007. Disponível em: <http://www.pnas.org/content/104/1/36.short>. Citado 2 vezes nas páginas 114 e 123.

FOTHERINGHAME, D.; BADDELEY, R. Nonlinear principal components analysis of neuronal spike train data. *Biological cybernetics*, v. 77, n. 4, p. 283–8, out. 1997. ISSN 0340-1200. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/9394446>. Citado na página 31.

FRéGNAC, Y. Hebbian Cell Assemblies. *Encyclopedia of Cognitive Science*, 2003. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/0470018860.s00355/full>. Citado na página 29.

FRIES, P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, v. 9, n. 10, p. 474–80, out. 2005. ISSN 1364-6613. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/16150631>. Citado na página 3.

GANSEL, K. S.; SINGER, W. Detecting multineuronal temporal patterns in parallel spike trains. *Frontiers in neuroinformatics*, v. 6, n. May, p. 18, jan. 2012. ISSN 1662-5196. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3357495&tool=pmcentrez&rendertype=abstract>. Citado 2 vezes nas páginas 9 e 23.

GERSTEIN, G. Neuronal assemblies. *IEEE Transactions on Biomedical Engineering*, v. 36, n. 1, p. 4 – 14, 1989. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=16444>. Citado na página 3.

GROSS, C. G. Genealogy of the "grandmother cell". *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, v. 8, p. 512–518, 2002. ISSN 1073-8584. Citado na página 6.

GRU, S.; GRüN, S. Data-driven significance estimation for precise spike correlation. *Journal of neurophysiology*, v. 101, n. January 2009, p. 1126–1140, 2009. ISSN 0022-3077. Citado 2 vezes nas páginas 16 e 23.

GUERRA, L.; ROBLES, V.; BIELZA, C.; LARRAAGA, P. A comparison of clustering quality indices using outliers and noise. *Intelligent Data Analysis*, v. 16, n. 4, p. 703–715, 2012. ISSN 1088467X. Citado na página 45.

GUHA, S.; RASTOGI, R.; SHIM, K. CURE: An Efficient Clustering Algorithm for Large Datasets. *Proceedings of the 1998 ACM SIGMOD international conference on Management of data - SIGMOD '98*, p. 73–84, 1998. Disponível em: <http://portal.acm.org/citation.cfm?doid=276304.276312>. Citado na página 123.

GUPTA, P. *Robust Clustering Algorithms*. Tese (Doutorado) — Georgia Institute of Technology, 2005. Citado 2 vezes nas páginas 116 e 123.

HARRIS, K. D. Neural signatures of cell assembly organization. *Nature reviews. Neuroscience*, v. 6, n. 5, p. 399–407, maio 2005. ISSN 1471-003X. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/15861182>. Citado 4 vezes nas páginas 3, 9, 15 e 133.

HAYKIN, S. *Neural Networks and Learning Machines*. Third. [s.n.], 1999. 906 p. ISBN 9780131471399. Disponível em: <http://www.amazon.com/Neural-Networks-Learning-Machines-Edition/dp/0131471392>. Citado 2 vezes nas páginas 2 e 30.

HEBB, D. *The Organization of Behavior*. [S.l.]: Wiley, New York, 1949. Citado na página 2.

HOCHBERG, L. R.; BACHER, D.; JAROSIEWICZ, B.; MASSE, N. Y.; SIMERAL, J. D.; VOGEL, J.; HADDADIN, S.; LIU, J.; CASH, S. S.; SMAGT, P. van der; DONOGHUE, J. P. *Reach and grasp by people with tetraplegia using a neurally controlled robotic arm*. 2012. 372–375 p. Disponível em: <http://www.nature.com/nature/journal/v485/n7398/abs/nature11076.html>. Citado na página 3.

HOCHBERG, L. R.; SERRUYA, M. D.; FRIEHS, G. M.; MUKAND, J. a.; SALEH, M.; CAPLAN, A. H.; BRANNER, A.; CHEN, D.; PENN, R. D.; DONOGHUE, J. P. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, v. 442, n. 7099, p. 164–71, jul. 2006. ISSN 1476-4687. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/16838014>. Citado na página 3.

HODGKIN, a. L.; HUXLEY, a. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Bulletin of Mathematical Biology*, v. 52, n. 1-2, p. 25–71, 1990. ISSN 00928240. Citado na página 13.

HOUGHTON, C.; VICTOR, J. D. Measuring representational distances – the spike train metrics approach. *Understanding Visual Population Codes – Towards a Common Multivariate Framework for Cell Recording and Functional Imaging*, p. in press, 2010. Disponível em: <http://www.maths.tcd.ie/~mnl/papers/tcd-mnl-7.pdf>. Citado 4 vezes nas páginas 39, 111, 126 e 133.

HUMPHRIES, M. D. Spike-train communities: finding groups of similar spike trains. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, v. 31, n. 6, p. 2321–36, fev. 2011. ISSN 1529-2401. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/21307268>. Citado 7 vezes nas páginas 9, 17, 18, 21, 39, 113 e 114.

HYVäRINEN, a.; OJA, E. Independent component analysis: Algorithms and applications. *Neural Networks*, v. 13, p. 411–430, 2000. ISSN 08936080. Citado na página 33.

IFFT, P. J.; SHOKUR, S.; LI, Z.; LEBEDEV, M. a.; NICOLELIS, M. a. L. A brain-machine interface enables bimanual arm movements in monkeys. *Science translational medicine*, v. 5, p. 210ra154, 2013. ISSN 1946-6242. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/24197735>. Citado 2 vezes nas páginas 3 e 129.

IZHIKEVICH, E. *Dynamical systems in neuroscience*. [s.n.], 2007. ISBN 9780262090438. Disponível em: <http://books.google.com/books?hl=en&lr=&id=kVjM6DFk-twC&oi=fnd&pg=PR15&dq=Dynamical+Systems+in+Neuroscience&ots=KTwvkZh4rl&sig=vqLwVq-BRMAPKNWxBkhzjm5HjFY>. Citado na página 13.

JACOBS, A. L.; FRIDMAN, G.; DOUGLAS, R. M.; ALAM, N. M.; LATHAM, P. E.; PRUSKY, G. T.; NIRENBERG, S. Ruling out and ruling in neural codes. *Proceedings of the National Academy of Sciences of the United States of America*, v. 106, n. 14, p. 5936–5941, 2009. ISSN 0027-8424. Citado na página 126.

KANDEL, E. R. *Priciples of neural science*. [S.l.]: McGraw-Hill Medical, 2012. ISBN 9780071810012. Citado na página 5.

KREUZ, T. Measures of spike train synchrony. *Scholarpedia*, v. 6, n. 10, p. 11934, out. 2011. ISSN 1941-6016. Citado 2 vezes nas páginas 24 e 26.

KREUZ, T.; CHICHARRO, D.; ANDRZEJAK, R. Measuring multiple spike train synchrony. *Journal of neuroscience ...* , p. 1–15, 2009. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0165027009003616>. Citado na página 9.

KREUZ, T.; CHICHARRO, D.; GRESCHNER, M.; ANDRZEJAK, R. G. Time-resolved and time-scale adaptive measures of spike train synchrony. *Journal of neuroscience methods*, Elsevier B.V., v. 195, n. 1, p. 92–106, jan. 2011. ISSN 1872-678X. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/21129402>. Citado 2 vezes nas páginas 25 e 133.

KREUZ, T.; CHICHARRO, D.; HOUGHTON, C.; ANDRZEJAK, R. G.; MORMANN, F. Monitoring spike train synchrony. *Journal of neurophysiology*, v. 109, n. 5, p. 1457–72, mar. 2013. ISSN 1522-1598. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/23221419>. Citado 2 vezes nas páginas 25 e 27.

KREUZ, T.; HAAS, J. S.; MORELLI, A.; ABARBANEL, H. D. I.; POLITI, A. Measuring spike train synchrony. *Journal of neuroscience methods*, v. 165, n. 1, p. 151–61, set. 2007. ISSN 0165-0270. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/17628690>. Citado 9 vezes nas páginas 9, 10, 17, 22, 24, 28, 40, 45 e 115.

KRUMIN, M.; SHOHAM, S. Generation of spike trains with controlled auto- and cross-correlation functions. *Neural computation*, v. 21, n. 6, p. 1642–64, jun. 2009. ISSN 0899-7667. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/19191596>. Citado 2 vezes nas páginas 127 e 133.

KRUPA, D. J.; MATELL, M. S.; BRISBEN, a. J.; OLIVEIRA, L. M.; NICOLELIS, M. a. Behavioral properties of the trigeminal somatosensory system in rats performing whisker-dependent tactile discriminations. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, v. 21, n. 15, p. 5752–5763, 2001. ISSN 1529-2401. Citado na página 53.

KRUPA, D. J.; WIEST, M. C.; SHULER, M. G.; LAUBACH, M.; NICOLELIS, M. a. L. Layer-specific somatosensory cortical activation during active tactile discrimination. *Science (New York, N.Y.)*, v. 304, n. 5679, p. 1989–1992, 2004. ISSN 0036-8075. Citado na página 53.

KRUSKAL, P. B.; STANIS, J. J.; MCNAUGHTON, B. L.; THOMAS, P. J. A binless correlation measure reduces the variability of memory reactivation estimates. In: *Statistics in Medicine*. [S.l.: s.n.], 2007. v. 26, n. April, p. 3997–4008. ISBN 0277-6715 (Print)\r0277-6715 (Linking). ISSN 02776715. Citado 3 vezes nas páginas 17, 23 e 113.

LAUBACH, M.; SHULER, M.; NICOLELIS, M. A. Independent component analyses for quantifying neuronal ensemble interactions. *Journal of neuroscience methods*, v. 94, n. 1, p. 141–54, dez. 1999. ISSN 0165-0270. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/10638821>. Citado 4 vezes nas páginas 9, 15, 18 e 35.

LEICHT, E. a.; LEICHT, E. a.; NEWMAN, M. E. J.; NEWMAN, M. E. J. Community structure in directed networks. *Physical Review Letters*, v. 100, p. 10–16, 2007. ISSN 0031-9007. Citado 2 vezes nas páginas 111 e 114.

LEWICKI, M. S. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network*, v. 9, n. 4, p. R53–R78, 1998. ISSN 0954-898X. Citado na página 134.

LIU, X.; MURATA, T.; WAKITA, K. Extending modularity by capturing the similarity attraction feature in the null model. *arXiv preprint arXiv:1210.4007*, out. 2012. Disponível em: <http://arxiv.org/abs/1210.4007>. Citado na página 114.

LIU, Y.; LI, Z.; XIONG, H.; GAO, X.; WU, J. Understanding of Internal Clustering Validation Measures. p. 1–6, 2013. Disponível em: <papers2://publication/uuid/30395C28-1BAB-4CEC-B569-9358E3006C1A>. Citado na página 45.

LONG, J. D.; CARMENA, J. M. A statistical description of neural ensemble dynamics. *Frontiers in computational neuroscience*, v. 5, n. November, p. 52, jan. 2011. ISSN 1662-5188. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3226070&tool=pmcentrez&rendertype=abstract>. Citado 3 vezes nas páginas 9, 29 e 114.

MACKE, J. H.; BERENS, P.; ECKER, A. S.; TOLIAS, A. S.; BETHGE, M. Generating spike trains with specified correlation coefficients. *Neural computation*, v. 21, n. 2, p. 397–423, fev. 2009. ISSN 0899-7667. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/19196233>. Citado 12 vezes nas páginas xxi, xxiii, xxv, 16, 17, 69, 73, 79, 80, 96, 126 e 127.

MACLEOD, K.; BäCKER, A.; LAURENT, G. Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature*, v. 395, n. October, p. 693–698, 1998. ISSN 0028-0836. Citado na página 23.

MALSBURG, C. von der; SCHNEIDER, W. A neural cocktail-party processor. *Biological cybernetics*, v. 54, p. 29–40, 1986. ISSN 0340-1200. Citado na página 8.

MARCENKO, V.; PASTUR, L. Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, v. 457, 1967. Disponível em: <http://www.turpion.org/php/full/infoFT.phtml?journal_id=sm&paper_id=1994>. Citado 3 vezes nas páginas 31, 117 e 118.

MARENCO, L. N.; WANG, R.; MORSE, T. M.; MILLER, P. L.; SHEPHERD, G. M. *NeuronDB - Home*. 2015. Disponível em: <https://senselab.med.yale.edu/neurondb/>. Citado na página 13.

MEILA, M. Comparing clusterings-an information based distance. *Journal of Multivariate Analysis*, v. 98, p. 873–895, 2007. ISSN 0047259X. Citado na página 49.

MILLER, E. K.; COHEN, J. D. An Integrative Theory of Prefrontal Cortex Function. p. 167–202, 2001. Citado na página 54.

MULANSKY, M.; BOZANIC, N.; SBURLEA, A.; KREUZ, T.; FIORENTINO, S. A guide to time-resolved and parameter-free measures of spike train synchrony. n. 1, p. 1–8, 2015. Citado 3 vezes nas páginas 28, 40 e 134.

NEWMAN, M. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 2006. Disponível em: <http://journals.aps.org/pre/abstract/10.1103/PhysRevE.74.036104>. Citado 5 vezes nas páginas 37, 38, 39, 111 e 114.

NICOLELIS, M. A.; FANSELOW, E. E.; GHAZANFAR, A. a. Hebb's Dream: The Resurgence of Cell Assemblies. *Neuron*, v. 19, n. 2, p. 219–221, ago. 1997. ISSN 08966273. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0896627300809320>. Citado 4 vezes nas páginas 3, 4, 112 e 133.

NICOSIA, V.; MANGIONI, G.; CARCHIOLO, V.; MALGERI, M. Extending the definition of modularity to directed graphs with overlapping communities. p. 22, 2008. ISSN 1742-5468. Disponível em: <http://arxiv.org/abs/0801.1647>. Citado na página 114.

OWEISS, K. *Statistical signal processing for neuroscience and neurotechnology*. [s.n.], 2010. ISBN 9780123750273. Disponível em: <http://books.google.com/books?hl=en&lr=&id=YqIo8_fsG0kC&oi=fnd&pg=PP2&dq=Statistical+Signal+Processing+for+Neuroscience+and+Neurotechnology&ots=Zo2AUKCn9d&sig=vXkCgDN_tMvk9qfdxTgvat78ZL4http://books.google.com/books?hl=en&lr=&id=YqIo8_fsG0kC&oi=fnd&pg=PP2&dq=Statistical+signal+processing+for+neuroscience+and+neurotechnology&ots=Zo2AULCo7e&sig=wjEtxBrwSJ7M_MfJVyWGy5qrELM>. Citado na página 127.

PAIS-VIEIRA, M.; CHIUFFA, G.; LEBEDEV, M.; YADAV, A.; NICOLELIS, M. a. L. Building an organic computing device with multiple interconnected brains. *Scientific Reports*, Nature Publishing Group, v. 5, p. 11869, 2015. ISSN 2045-2322. Disponível em: <http://www.nature.com/doifinder/10.1038/srep11869>. Citado na página 129.

PAIS-VIEIRA, M.; LEBEDEV, M.; KUNICKI, C.; WANG, J.; NICOLELIS, M. A. L. A brain-to-brain interface for real-time sharing of sensorimotor information. *Scientific reports*, v. 3, p. 1319, jan. 2013. ISSN 2045-2322. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3584574&tool=pmcentrez&rendertype=abstract>. Citado na página 129.

PAIS-VIEIRA, M.; LEBEDEV, M. A.; WIEST, M. C.; NICOLELIS, M. A. L. Simultaneous Top-down Modulation of the Primary Somatosensory Cortex and Thalamic Nuclei during

Active Tactile Discrimination. *Journal of Neuroscience*, v. 33, n. 9, p. 4076–4093, 2013. ISSN 0270-6474. Disponível em: <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1659-12.2013>. Citado 4 vezes nas páginas 4, 53, 113 e 127.

PAIVA, A.; PARK, I.; PRÍNCIPE, J. A comparison of binless spike train measures. *Neural Computing and Applications*, n. 3, p. 405–419, 2010. Disponível em: <http://link.springer.com/article/10.1007/s00521-009-0307-6>. Citado 8 vezes nas páginas 8, 17, 20, 27, 110, 111, 113 e 127.

PANZERI, S.; BRUNEL, N.; LOGOTHETIS, N. K.; KAYSER, C. Sensory neural codes using multiplexed temporal scales. *Trends in neurosciences*, v. 33, n. 3, p. 111–20, mar. 2010. ISSN 1878-108X. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/20045201>. Citado 5 vezes nas páginas 8, 9, 23, 28 e 134.

PAPOULIS, A. *Probability, Random Variables and Stochastic Processes*. 3rd. ed. [S.l.]: McGrawHill, 1991. 666 p. ISBN 9780071008709. Citado 2 vezes nas páginas 33 e 109.

PAXINOS, G.; WATSON, C. *The Rat Brain in Stereotaxic Coordinates*. 6th. ed. [S.l.]: Elsevier, 2007. ISBN 9780125476126. Citado na página 55.

PEYRACHE, A.; BENCHENANE, K.; KHAMASSI, M.; WIENER, S. I.; BATTAGLIA, F. P. Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution. *Journal of computational neuroscience*, v. 29, n. 1-2, p. 309–25, ago. 2010. ISSN 1573-6873. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2940043&tool=pmcentrez&rendertype=abstract>. Citado 3 vezes nas páginas 8, 18 e 31.

PICADO-MUINO, D.; BORGELT, C.; BERGER, D.; GERSTEIN, G.; GRüN, S. Finding neural assemblies with frequent item set mining. *Frontiers in neuroinformatics*, v. 7, n. May, p. 9, 2013. ISSN 1662-5196. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3668274&tool=pmcentrez&rendertype=abstract>. Citado na página 9.

PIPA, G. The neuronal code. n. August, 2006. Disponível em: <http://www.neuroxidence.com/PhD_Gordon_Pipa.pdf>. Citado na página 8.

POLIKOV, V. S.; TRESCO, P. a.; REICHERT, W. M. Response of brain tissue to chronically implanted neural electrodes. *Journal of Neuroscience Methods*, v. 148, n. 1, p. 1–18, 2005. ISSN 01650270. Citado na página 134.

QIN, W.; YU, C. Neural pathways conveying novisual information to the visual cortex. *Neural plasticity*, v. 2013, p. 864920, jan. 2013. ISSN 1687-5443. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3690246&tool=pmcentrez&rendertype=abstract>. Citado na página 3.

QUIANQUIROGA, R.; PANZERI, S. Extracting information from neuronal populations: information theory and decoding approaches. *Nature reviews. Neuroscience*, v. 10, n. 3, p. 173–85, mar. 2009. ISSN 1471-0048. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/19229240>. Citado 2 vezes nas páginas 2 e 9.

QUIROGA, R. Q.; KREUZ, T.; GRASSBERGER, P. Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, v. 66, p. 1–9, 2002. ISSN 15393755. Citado na página 23.

RAIMONDO, F.; KAMIENKOWSKI, J. E.; SIGMAN, M.; Fernandez Slezak, D. CUDAICA: GPU optimization of infomax-ICA EEG analysis. *Computational Intelligence and Neuroscience*, v. 2012, 2012. ISSN 16875265. Citado na página 113.

RAMAKRISHNAN, A.; IFFT, P. J.; PAIS-VIEIRA, M.; Woo Byun, Y.; ZHUANG, K. Z.; LEBEDEV, M. a.; NICOLELIS, M. a.L. Computing Arm Movements with a Monkey Brainet. *Scientific Reports*, Nature Publishing Group, v. 5, p. 10767, 2015. ISSN 2045-2322. Disponível em: <http://www.nature.com/doifinder/10.1038/srep10767>. Citado na página 129.

REEP, R. L.; CHANDLER, H. C.; KING, V.; CORWIN, J. V. Rat posterior parietal cortex: topography of corticocortical and thalamic connections. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, v. 100, n. 1, p. 67–84, 1994. ISSN 0014-4819. Citado na página 54.

RENDóN, E.; ABUNDEZ, I.; ARIZMENDI, A.; QUIROZ, E. M. Internal versus External cluster validation indexes. *International Journal of Computers and Communications*, v. 5, n. 1, p. 27—-34, 2011. Disponível em: <http://w.naun.org/multimedia/UPress/cc/20-463.pdf>. Citado na página 45.

RIEKE, F.; WARLAND, D.; De Ruyter Van Steveninck, R.; BIALEK, W.; STEVENINCK, R. R. van; BIALEK, W. *Spikes: Exploring the Neural Code*. [S.l.]: Bradford Book - MIT Press, Cambridge MA, 1997. xvi, 395 p. p. ISSN 0028-0836. ISBN 0262181746. Citado na página 8.

ROMANSKI, L. M. Convergence of Auditory, Visual, and Somatosensory Information in Ventral Prefrontal Cortex. In: MURRAY, M.; WALLACE, M. (Ed.). *The Neural Bases of Multisensory Processes.* CRC Press, 2011. p. 667–682. ISBN 9781439812198. Disponível em: <https://books.google.com.br/books?id=-WvMBQAAQBAJ>. Citado na página 54.

ROSSANT, C.; GOODMAN, D. F. M.; FONTAINE, B.; PLATKIEWICZ, J.; MAGNUSSON, A. K.; BRETTE, R. Fitting neuron models to spike trains. *Frontiers in Neuroscience*, v. 5, n. February, p. 1–8, 2011. ISSN 16624548. Citado na página 113.

ROSSANT, C.; GOODMAN, D. F. M.; PLATKIEWICZ, J.; BRETTE, R. Automatic fitting of spiking neuron models to electrophysiological recordings. *Frontiers in neuroinformatics*, v. 4, n. March, p. 2, 2010. ISSN 1662-5196. Citado na página 113.

ROSSUM, M. C. van. A novel spike distance. *Neural computation*, v. 13, p. 751–763, 2001. ISSN 0899-7667. Citado na página 21.

ROUSSEEUW, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, v. 20, p. 53–65, 1987. ISSN 03770427. Citado na página 46.

SABATTINI, R. *Neurons and Synapses: The History of Its Discovery*. 2003. Disponível em: <http://www.cerebromente.org.br/n17/history/neurons2_i.htm>. Citado na página 2.

SAKURAI, Y. The search for cell assemblies in the working brain. *Behavioural brain research*, v. 91, n. 1-2, p. 1–13, mar. 1998. ISSN 0166-4328. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/9578434>. Citado 4 vezes nas páginas 2, 3, 7 e 14.

SANTOS, V. Lopes-dos; CONDE-OCAZIONEZ, S.; NICOLELIS, M. A. L.; RIBEIRO, S. T.; TORT, A. B. L. Neuronal assembly detection and cell membership specification by principal component analysis. *PloS one*, v. 6, n. 6, jan. 2011. ISSN 1932-6203. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3115970&tool=pmcentrez&rendertype=abstract>. Citado 5 vezes nas páginas 9, 18, 31, 32 e 119.

SANTOS, V. Lopes-dos; RIBEIRO, S.; TORT, A. B. L. Detecting cell assemblies in large neuronal populations. *Journal of neuroscience methods*, Elsevier B.V., v. 220, n. 2, p. 149–66, nov. 2013. ISSN 1872-678X. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/23639919>. Citado 8 vezes nas páginas 9, 14, 18, 31, 32, 33, 35 e 36.

SCHöLKOPF, B.; SMOLA, A.; MüLLER, K. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, n. 44, 1998. Disponível em: <http://www.mitpressjournals.org/doi/abs/10.1162/089976698300017467>. Citado na página 31.

SCHREIBER, S.; FELLOUS, J.; WHITMER, D.; TIESINGA, P.; SEJNOWSKI, T. A new correlation-based measure of spike timing reliability. *Neurocomputing*, v. 52-54, p. 925–931, 2003. ISSN 09252312. Citado na página 21.

SCHWARZ, D. A.; LEBEDEV, M. A.; HANSON, T. L.; DIMITROV, D. F.; LEHEW, G.; MELOY, J.; RAJANGAM, S.; SUBRAMANIAN, V.; IFFT, P. J.; LI, Z.; RAMAKRISHNAN, A.; TATE, A.; ZHUANG, K. Z.; NICOLELIS, M. A. L. Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys. *Nature methods*, v. 11, n. 6, abr. 2014. ISSN 1548-7105. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/24776634>. Citado 3 vezes nas páginas 4, 112 e 134.

SEJNOWSKI, T.; CHATTARJI, S.; STANTON, P. Induction of synaptic plasticity by Hebbian covariance in the hippocampus. *The computing neuron*, 1989. Disponível em: <http://dl.acm.org/citation.cfm?id=103945>. Citado na página 29.

SERRUYA, M. D.; HATSOPOULOS, N. G.; PANINSKI, L.; FELLOWS, M. R.; DONOGHUE, J. P. *Brain-machine interface: Instant neural control of a movement signal*. 2002. 141–142 p. Citado na página 3.

SHERMAN, S. M. What is the Function of the Thalamus? In: *23 Problems in Systems Neuroscience*. [s.n.], 2006. p. 65–83. Disponível em: <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195148220.001.0001/acprof-9780195148220-chapter-4>. Citado 2 vezes nas páginas 109 e 110.

SHOKUR, S.; O'DOHERTY, J. E.; WINANS, J. a.; BLEULER, H.; LEBEDEV, M. a.; NICOLELIS, M. a. L. Expanding the primate body schema in sensorimotor cortex by virtual touches of an avatar. *Proceedings of the National Academy of Sciences of the United States of America*, v. 110, n. 37, p. 15121–6, 2013. ISSN 1091-6490. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3773736&tool=pmcentrez&rendertype=abstract>. Citado 2 vezes nas páginas 3 e 129.

SIEBEN, K.; RöDER, B.; HANGANU-OPATZ, I. L. Oscillatory entrainment of primary somatosensory cortex encodes visual control of tactile processing. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, v. 33, n. 13, p. 5736–49, 2013. ISSN 1529-2401. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/23536087>. Citado na página 54.

SOLO, V.; PASHA, S. A. Point-Process Principal Components Analysis via Geometric Optimization. *Neural Computation*, v. 122, p. 1–22, 2012. ISSN 0899-7667. Citado na página 114.

STERIADE, M.; MCCORMICK, D.; SEJNOWSKI, T. Thalamocortical oscillations in the sleeping and aroused brain. *Science*, v. 262, n. 5134, p. 679–685, 1993. Disponível em: <http://www.sciencemag.org/content/262/5134/679.short>. Citado na página 3.

STILLA, R.; DESHPANDE, G.; LACONTE, S.; HU, X.; SATHIAN, K. Posteromedial parietal cortical activity and inputs predict tactile spatial acuity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, v. 27, n. 41, p. 11091–11102, 2007. ISSN 0270-6474. Citado na página 54.

VARELA, F.; LACHAUX, J.; RODRIGUEZ, E.; MARTINERIE, J. The brainweb: phase synchronization and large-scale integration. *Nature reviews ...* , v. 2, n. April, 2001. Disponível em: <http://www.nature.com/nrn/journal/v2/n4/abs/nrn0401_229a.html>. Citado na página 3.

VASCONCELOS, N.; PANTOJA, J.; BELCHIOR, H.; CAIXETA, F. V.; FABER, J.; FREIRE, M. A. M.; COTA, V. R.; Anibal de Macedo, E.; LAPLAGNE, D. a.; GOMES, H. M.; RIBEIRO, S. Cross-modal responses in the primary visual cortex encode complex objects and correlate with tactile discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, v. 108, n. 37, p. 15408–15413, 2011. ISSN 0027-8424. Citado na página 54.

VICTOR, J. D. Spike Train Metrics. *Neurology*, v. 15, n. 5, p. 585–592, 2009. Citado na página 20.

VICTOR, J. D.; PURPURA, K. P. Nature and precision of temporal coding in visual cortex: a metric-space analysis. *Journal of neurophysiology*, v. 76, n. August, p. 1310–1326, 1996. ISSN 0022-3077. Citado na página 20.

VOEGTLIN, T. Recursive principal components analysis. *Neural networks : the official journal of the International Neural Network Society*, v. 18, n. 8, p. 1051–63, out. 2005. ISSN 0893-6080. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/16181769>. Citado na página 31.

WANG, Q.; ZHANG, H.; PERC, M.; CHEN, G. Multiple firing coherence resonances in excitatory and inhibitory coupled neurons. *Communications in Nonlinear Science*, p. 1–7, 2012. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1007570412000755>. Citado na página 29.

WESSBERG, J.; STAMBAUGH, C. R.; KRALIK, J. D.; BECK, P. D.; LAUBACH, M.; CHAPIN, J. K.; KIM, J.; BIGGS, S. J.; SRINIVASAN, M. a.; NICOLELIS, M. a. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, v. 408, n. 6810, p. 361–5, nov. 2000. ISSN 0028-0836. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/11099043>. Citado na página 3.

WHITLOCK, J. R.; SUTHERLAND, R. J.; WITTER, M. P.; MOSER, M.-B.; MOSER, E. I. Navigating from hippocampus to parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, v. 105, n. 39, p. 14755–14762, 2008. ISSN 0027-8424. Citado na página 54.

WIEST, M. C.; THOMSON, E.; PANTOJA, J.; NICOLELIS, M. a. L. Changes in S1 neural responses during tactile discrimination learning. *Journal of neurophysiology*, v. 104, n. 1, p. 300–312, 2010. ISSN 0022-3077. Citado na página 103.

WINTERS, B. D.; REID, J. M. A distributed cortical representation underlies crossmodal object recognition in rats. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, v. 30, n. 18, p. 6253–6261, 2010. ISSN 0270-6474. Citado na página 54.

XU, B.; GONG, Y. B.; WANG, L.; YANG, C. L. Dependence of delay-induced coherence resonance on time-periodic coupling strength in Newman-Watts neuronal networks. *The European Physical Journal B*, v. 85, n. 9, p. 299, set. 2012. ISSN 1434-6028. Disponível em: <http://www.springerlink.com/index/10.1140/epjb/e2012-30452-0>. Citado na página 29.

YANG, X. L.; SENTHILKUMAR, D. V.; KURTHS, J. Impact of connection delays on noise-induced spatiotemporal patterns in neuronal networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, v. 22, n. 4, dez. 2012. ISSN 1089-7682. Disponível em: <http://scitation.aip.org/content/aip/journal/chaos/22/4/10.1063/1.4772999>. Citado na página 29.

ZANGALADZE, A.; EPSTEIN, C. M.; GRAFTON, S. T.; SATHIAN, K. Involvement of visual cortex in tactile discrimination of orientation. *Nature*, v. 401, n. 6753, p. 587–590, 1999. ISSN 0028-0836. Citado na página 54.

ANEXO A – Produção técnica e acadêmica

- **Sardeto, C. D.** ; Kunicki, C. ; Brasil, F. ; Moioli, R. C. . *Mapeamento das limitações da análise de componentes principais como como método de detecção de assembleias neurais*. IX Reunião Regional da FeSBE, João Pessoa, 2014.

- **Sardeto, C. D.** ; Kunicki, C. ; Brasil, F. ; Moioli, R. C. . *Limitations of principal component analysis as a method to detect neuronal assemblies*. International Conference on E-health Networking, Application & Services, Natal, 2014.

- **Sardeto, C. D.** ; Kunicki, C. ; Brasil, F. ; Moioli, R. C. ; Nicolelis, M. . *Phase Locking Value Analysis in active tactile discrimination task*. Society for Neuroscience, Washington DC, 2014.

- Kunicki, C. ; Moioli, R. C. ; **Sardeto, C. D.** ; Morya, E. ; Nicolelis, M. . *Neuronal activity patterns of posterior parietal cortical activity during active tactile discrimination task*. Society for Neuroscience, Washington DC, 2014.

- Silva, M. I. ; Moioli, R. C. ; **Sardeto, C. D.** ; Morya, E. ; Kunicki, C. ; Nicolelis, M. . *Distributed representation of tactile information streams in prefrontal cortex*. Society for Neuroscience, Washington DC, 2014.