

# O uso de Mapas Auto-organizáveis como ferramenta de Análise Exploratória para Testes Cognitivos destinados a medir o Desempenho Escolar

Johne M. Jarske<sup>1</sup>, Alessandra Gotuso Seabra<sup>2</sup>, Leandro A. Silva<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Engenharia Elétrica e Computação

<sup>2</sup>Programa de Pós-Graduação em Distúrbios de Aprendizagem

Universidade Presbiteriana Mackenzie

Rua da Consolação 930 – 01302-907 – São Paulo – SP – Brasil

johnejarske@hotmail.com

alessandra.seabra, leandroaugusto.silva@mackenzie.br

**Abstract.** *This paper proposes the use of Self-organizing Maps, as an Exploratory Data Analysis tool for supporting findings and diagnoses on school performance, focusing on the analysis of cognitive tests capable to measure the development of interations between school performance, word recognition and reading.*

**Resumo.** *Este artigo propõe o uso de Mapas Auto-organizáveis (Self-organizing Maps), como ferramenta para Análise Exploratória de Dados (Exploratory Data Analysis), no apoio a descobertas e diagnósticos relativos ao desempenho escolar, com foco na análise de testes voltados para a mensuração do desenvolvimento das interações entre o desempenho escolar, reconhecimento de palavras e de leitura.*

## 1. Introdução

De acordo com [Dias et al. 2015], o campo de estudo das habilidades cognitivas considera como fator relevante para o desempenho acadêmico, a estratégia de aprendizagem adotada pelos estudantes, incluindo as capacidades de auto-organização e controle durante o processo de aprendizado. Estudos conduzidos por [Duncan et al. 2007] mostram que a capacidade de concentração diante do aprendizado de matemática e leitura desde o início do aprendizado escolar são indicadores eficientes do desempenho escolar, enquanto que outros fatores como o nível socioeconômico, gênero, aspectos sociais e emocionais não contribuem de forma significativa.

Em seu estudo, [Dias et al. 2015] identificaram padrões de correlação entre as estratégias de leitura, as habilidade de compreensão da leitura e da audição e o desempenho acadêmico em função dos anos de escolaridade, ressaltando as habilidades que tendem a ser mais proeminentes ou cruciais para o desempenho acadêmico em diferentes níveis de ensino. Nesse estudo foram analisados os resultados obtidos pela aplicação de dois testes, o Teste de Competência de Leitura de Palavras e Pseudopalavras - TCLPP e o Teste Contrastivo de Compreensão Auditiva e de Leitura - TCCAL, capazes de medir o

desenvolvimento cognitivo relacionados a leitura, especificamente aspectos relacionados ao reconhecimento de palavras, a audição e a leitura silenciosa.

Como é comum em pesquisas ligadas a área de ciências do comportamento, [Baldwin et al. 2014, Wilkinson 1999, Schmidt 1996, Heiman 2011, Dias et al. 2015], foram utilizadas técnicas de Inferência Estatística, em especial, a Análise de Variância (*Analysis of Variance*, ANOVA), Análise de Variância Multivariada (*Multivariate Analysis of Variance*, MANOVA), Correlação de Pearson (*Pearson's Correlation*) e o Teste de Bonferroni (*Bonferroni Correction*), para fundamentar a análise dos dados.

O presente artigo, propõe-se a revisar o estudo realizado por [Dias et al. 2015], adotando técnicas de Análise Exploratória de Dados (*Exploratory Data Analysis*, EDA), e em particular, o uso de Mapas Auto-organizáveis (*Self-organizing Maps*, SOM) na análise de dados de desempenho escolar.

Justifica-se o uso da EDA, neste trabalho, por esta reunir um conjunto de ferramentas de visualização de dados, descoberta de padrões, formulação, desenvolvimento e refinamento de hipóteses, que não são possíveis de se obter pelas técnicas de estatística inferencial [Behrens 1997]. Ademais, esta proposta está alinhada com a *American Psychological Association (APA)*, que recomenda combinar o uso de estatística inferencial com técnicas de visualização de dados objetivando melhorar a avaliação dos pressupostos e divulgação dos resultados [Wilkinson 1999].

Dentre as ferramentas de EDA, justifica-se a adoção do SOM devido aos potenciais benefícios que a ferramenta pode proporcionar, tais como: i) ser capaz de maximizar o conhecimento sobre dados com muitas dimensões (no caso, os diversos fatores que afetam o desempenho escolar), permitindo a visualização do espaço multidimensional num plano bidimensional [Kohonen 2013]; ii) ser capaz de agrupar elementos com características semelhantes; iii) realizar a ordenação topológica dos dados (os elementos que possuem características semelhantes são exibidos próximos entre si e os dissemelhantes mais afastados) [Kohonen 2013]; iv) proporcionar a visualização de variáveis não treinadas, possibilitando visualizar a sua distribuição topológica em função dos dados treinados; vi) ser capaz de identificar casos anômalos e *outliers* [Muñoz and Muruzábal 1998], para os quais ANOVA é extremamente sensível [Salibian-barrera et al. 2016]; vii) proporcionar visualizações capazes de revelar novas hipóteses sobre o conjunto de dados e a relação entre suas múltiplas variáveis.

Nas próximas sessões serão discutidos uso de SOM como ferramenta de EDA, será apresentado o TCLPP como teste cognitivo capaz de medir o desempenho na leitura, a metodologia empregada no estudo, bem como os resultados obtidos. Por fim, na última sessão, serão feitas as considerações finais e sugestões de pesquisa.

## 2. O uso de SOM como ferramenta de EDA

De acordo com [Good 1983], a análise dos dados pode ser dividida em dois tipos: i) Análise Confirmatória de Dados (*Confirmatory Data Analysis*, CDA) e ii) Análise Exploratória de Dados (*Exploratory Data Analysis*, EDA), sendo que a primeira preocupa-se com experimentos, teste de significância, estimativas e predições enquanto que a segunda preocupa-se primariamente com a formulação de hipóteses.

O objetivo central da EDA é apresentar os dados de uma forma compreensível,

e que, ao mesmo tempo, preserve a essência da informação dos dados originais. O algoritmo SOM, atende aos objetivos da EDA, uma vez que une a projeção de dados multidimensionais complexos numa estrutura mais simples (normalmente bidimensional) [Kaski and Kohonen 1996].

Proposto por Teuvo Kohonen em 1982, o SOM é um tipo de rede neural artificial interconectada e não supervisionada, que permite um mapeamento auto ajustável do espaço de dados multidimensionais. Originalmente o SOM foi pensado para resolver problemas envolvendo tarefas como agrupamento de dados, visualização e abstração, mas com o tempo, foram sendo descobertas novas aplicações em diferentes áreas.

O SOM pode ser utilizado para um estudo mais amplo da correlação entre as múltiplas variáveis existentes de um fenômeno, sem restrição prévia do número de variáveis a serem analisadas, devido a sua propriedade de mapear dados com elevado número de dimensões em um espaço dimensional menor, ou seja, convertendo relações estatísticas não-lineares complexas, em relações geométricas simples em um mapa de dimensões reduzidas, mantendo a relação topológica original, na medida em que a localização física dos dados no mapa mostra a similaridade relativa entre os mesmos no espaço multidimensional original [Kohonen 2013].

O SOM, em geral, é representado por uma matriz (*array*) bidimensional de nodos, denominado mapa de características (*feature map*) e convencionalmente representado por um arranjo geométrico, geralmente quadrangular, de nodos interconectados, que implementam uma rede competitiva do tipo "winner-take-all" *network* [Harp et al. 1995].

No caso da representação bidimensional, a mais comum, o treinamento dos dados multidimensionais de entrada, resultam numa matriz pré-definida, de x colunas e y linhas. Cada nodo da matriz é caracterizada pela sua posição  $\{i, j\}$  e por um vetor p-dimensional denominado vetor de características (*codebook vector*).

O vetor de características de cada nodo é comparado com cada novo conjunto de valores de entrada, computando-se as distâncias euclidianas entre estes para estabelecer qual nodo é mais próximo dos valores de entrada corrente. Este nodo, torna-se então o vencedor, de forma que os valores de entrada próximos uns dos outros ficarão próximos na representação bidimensional proporcionada pelo algoritmo.

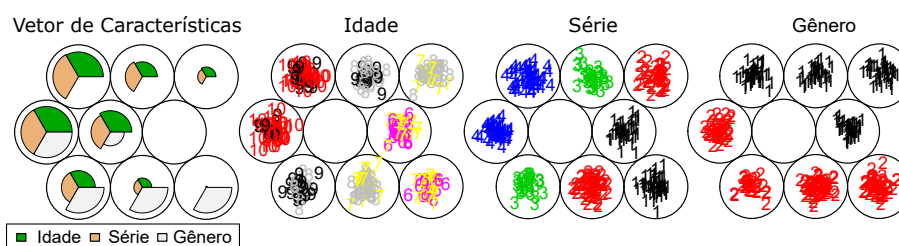
A Figura 1, apresenta um conjunto de quatro mapas resultantes do treinamento do SOM com os atributos idade, série e gênero de uma base de dados estudantil. Os mapas são formados por uma rede de quadrangular de nove nodos, sendo que cada nodo corresponde a um neurônio da rede neural.

No primeiro mapa da Figura 1, denominado Vetor de Características, é plotado o vetor resultante do treinamento do SOM na forma de elementos circulares coloridos. Nos três mapas seguintes, foram plotados, respectivamente, os valores dos atributos idade, série e gênero, nos respectivos nodos onde estes atributos estão representados. Pela análise dos mapas, é possível observar a ordenação topológica do espaço amostral da base de dados, onde os valores mais semelhantes foram dispostos próximos entre si enquanto os mais dissemelhantes ficaram mais afastados.

Por conta desta distribuição topológica, os valores mais altos para idade (nove e dez anos) e série escolar (4ª série) estão associados aos primeiros nodos da esquerda na

primeira e segunda linhas do mapa, sendo que no nodo da primeira linha concentram-se apenas os meninos (1) e no da segunda linha as meninas (2). Nos últimos nodos da direita, na segunda e terceira linhas, estão associados os menores valores para idade (seis e sete anos) e série (1ª série), sendo que no nodo da segunda linha concentram-se apenas os meninos (1) e no da segunda linha as meninas (2). Valores intermediários estão localizados nos demais nodos, os mais similares próximos entre si.

O nodo central dos mapas da Figura 1 (nodo vazio no segundo, terceiro e quarto mapas), não está associado a nenhum valor de amostra, no entanto, o vetor de características do nodo possui valores diferentes de zero (primeiro mapa).



**Figura 1. Rede SOM treinada com informações sobre idade, série e gênero.**

Como poderá ser observado nos próximos tópicos, as propriedades de visualização do SOM, o tornam uma ferramenta de dados potencialmente útil para a análise exploratória de dados educacionais, especialmente na exploração de informações pouco estudadas, com objetivo de ajudar a preencher lacunas ou ainda para a descoberta de novas informações, particularmente, de dados multidimensionais.

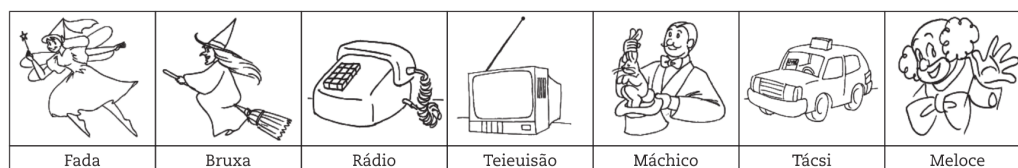
### 3. Teste de Competência de Leitura de Palavras e Pseudopalavras

O TCLPP é um instrumento psicométrico e neuropsicológico cognitivo para avaliar a competência da leitura silenciosa de palavras isoladas e servir de coadjuvante para diagnóstico diferencial de distúrbios de aquisição de leitura [Capovilla and Seabra 2010].

De acordo com [Dias and Seabra 2014], o TCLPP contém 70 itens de teste, divididos em sete subtestes, contendo 10 itens de teste cada. Cada item de teste é composto por um par de figura e palavra.

Os subtestes do TCLPP são: palavras corretas regulares - CR, palavras corretas irregulares - CI, palavras com incorreções ou trocas semânticas - TS, pseudopalavras com trocas visuais - TV, pseudopalavras com trocas fonológicas - TF, pseudopalavras homófonas - PH e pseudopalavras estranhas - PE. A Figura 2, exemplifica cada um dos tipos de pares de testes do TCLPP [Dias and Seabra 2014].

Os 70 itens de teste são apresentados às crianças, que precisam responder se a palavra está correta ou errada. Desta forma, o TCLPP permite interpretar os dados do padrão de leitura específico apresentado pela criança quanto ao modelo do desenvolvimento de leitura e escrita, e inferir o estágio de desenvolvimento (logográfico, alfabético, ortográfico) em que ela se encontra e as estratégias de leitura (ideovisual ou logográfica, perilexical ou fonológica, lexical) prevalentes. A cada resposta correta é somado 0,1 ponto ao subteste correspondente, de forma que a nota de cada subteste pode variar entre 0 e 1 pontos [Capovilla and Seabra 2010].



**Figura 2. Exemplos de itens do TCLPP. Na ordem de apresentação os pares de palavras e figuras: CR: FADA - fada; CI: BRUXA - bruxa; TS: RÁDIO - telefone; TV: TEIEUISÃO - televisão; TF: MÁCHICO - mágico; PH: TÁCSI - táxi e PE: MELOCE - palhaço [Dias and Seabra 2014, pp. 239].**

O padrão de distribuição dos tipos de erros indicam as estratégias de leitura utilizadas, de forma que, o padrão de resposta ao teste pode auxiliar na identificação da dificuldade de leitura de uma determinada criança [Dias and Seabra 2014].

#### 4. Metodologia

O presente artigo propõe o uso de EDA, mais particularmente, a utilização do SOM, para a análise de dados resultantes da aplicação de testes voltados para avaliar o desenvolvimento cognitivos dos estudantes.

Tradicionalmente, estes dados são analisados com emprego de técnicas de análise estatística, como ANOVA, MANOVA, Coeficiente de Pearson, Teste de Bonferroni, cujas vantagens e limitações como técnicas de análise são conhecidas e bem documentadas na literatura [Hoaglin 2003].

Os dados utilizados foram coletados por pesquisadores ligados ao estudo da psicologia educacional, numa escola pública da área metropolitana da cidade de São Paulo, contendo 443 amostras, envolvendo meninos e meninas entre 6 e 14 anos, matriculados entre a primeira e quarta séries do ensino fundamental. Desta base de dados foram extraídos os sete atributos relacionados ao teste TCLPP, além de informações referentes a idade, série escolar e gênero das crianças envolvidas nos testes. Esta base de dados, é uma versão ampliada da base original utilizada no artigo de [Dias et al. 2015].

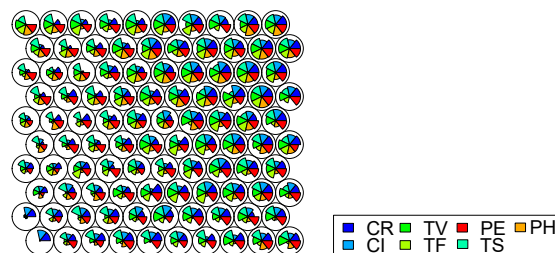
Após seleção dos dados de interesse da pesquisa, a base de dados foi pré-processada, atendendo-se os mesmos critérios definidos por [Dias et al. 2015], tendo sido excluídos os dados relativos às crianças que não atendessem a determinados requisitos de idade e série-escolar. Os critérios adotados foram: i) 1ª série - crianças entre 6 e 7 anos; ii) 2ª série - crianças entre 7 e 8 anos; iii) 3ª série - crianças entre 8 e 9 anos; iv) 4ª série - crianças entre 9 e 10 anos. Amostras contendo dados incompletos também foram removidos da base de conhecimento. Todos os atributos referentes ao teste TCLPP variam numa escala de valor entre 0 e 1, não tendo sido necessária a normalização dos valores.

A amostragem final resultou em um conjunto de dados contendo 338 crianças, dos quais 160 meninos e 178 meninas, cursando entre a 1ª e 4ª séries do ensino fundamental, com idade entre 6 e 10 anos. Deste total, i) 87 estudantes, 38 meninos e 49 meninas da 1ª série; ii) 99 estudantes, 43 meninos e 56 meninas da 2ª série; iii) 67 estudantes, meninos e 30 meninas da 3ª série; iv) 85 estudantes, 42 meninos e 43 meninas da 4ª série.

Na próxima seção serão analisadas um conjunto de imagens resultantes do treinamento do SOM com os dados da base de conhecimento obtida, com o propósito de demonstrar as propriedades de visualização proporcionados pela ferramenta.

## 5. Análise dos resultados

A Figura 3, exibe o mapa de características resultante do treinamento do SOM com os subtestes do TCLPP. A impressão dos valores do vetor de características, correspondentes a cada subteste, na forma de elementos circulares, nos permite observar peso de cada subteste na composição no nodo. Pode-se observar que o peso (representado pela área do elemento circular correspondente cada subteste) para os subtestes PE e TS são relativamente homogêneos por todo o mapa, enquanto que para os subtestes CR e CI, os pesos são menores na parte esquerda e maiores na parte direita do mapa.



**Figura 3. Mapa de características dos sete subtestes do TCLPP, representados na forma de elementos circulares coloridos.**

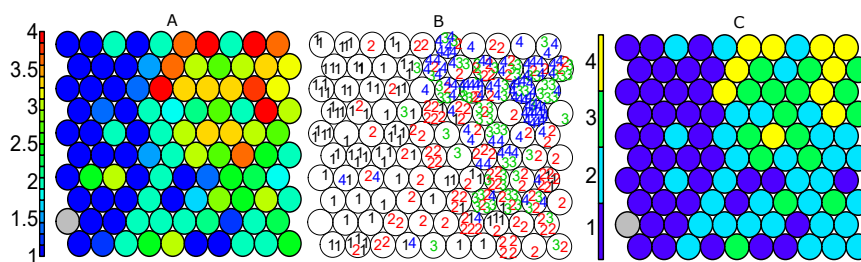
A Figura 4, exibe três visões diferentes do mapa de características sobrepostos com a série escolar dos estudantes (um atributo categórico, que não participou do treinamento do SOM). Na Figura 4-A, a série escolar é representada na forma de um mapa de calor, com as cores variando entre tons azul, verde, amarelo, vermelho e cinza, representando respectivamente a 1ª, 2ª, 3ª e 4ª séries. Na Figura 4-B, os dados a série escolar são sobrepostos sobre o mapa de características. Na Figura 4-C, a série escolar dominante em cada nodo é destacada sobre o mapa de características. O nodo que aparece com a cor cinza nas Figuras 4-A e 4-C não contém nenhum elemento do conjunto de entrada, o que pode ser confirmado na Figura 4-C.

As visões apresentadas na Figura 4, auxiliam na identificação de possíveis *outliers*, como os estudantes da 4ª série, que aparecem na Figura 4-B, numa região onde são esperados estudantes da 1ª série.

O nodo de cor cinza, que aparece nas Figuras 4-A e 4-C, segundo nodo do lado esquerdo, contando de baixo para cima do mapa, não contém nenhum exemplar e por isto aparecem representados nesta cor.

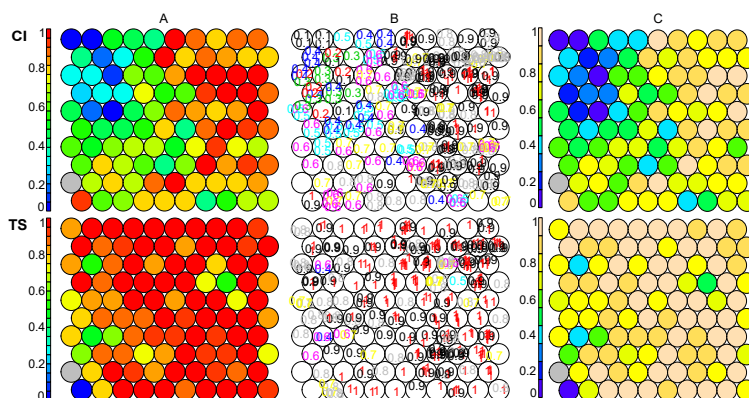
Comparando-se a Figura 3, com a Figura 4-A, pode-se verificar uma correlação positiva entre o TCLPP e a série escolar, pois, à medida que a pontuação dos subtestes do TCLPP aumentam, a classe escolar também aumenta. O mapa 4-B, revela uma separação bastante nítida dos estudantes da 1ª série, que se concentram na parte esquerda do mapa onde estão localizadas as menores pontuações do TCLPP e uma separação menos nítida entre as crianças da 2ª série, que se concentram na parte central inferior do mapa, e a 3ª e 4ª séries que se apresentam misturados na parte superior direita do mapa, confirmando os estudos de [Dias et al. 2015], que afirma que a pontuação do TCLPP aumenta à medida que a série escolar aumenta e que não há diferenciação significativa entre essas duas séries.

A Figura 5, exibe um mapa de características com a distribuição dos valores dos atributos CI (palavras corretas irregulares) e TS (troca semântica), sendo que: CI A -



**Figura 4. A - mapa de calor da série escolar (valores médios da série escolar); B - dados da série escolar sobrepostos ao mapa; C - os valores predominantes (moda) (moda) da série escolar.**

apresenta o mapa de calor para o subtteste CI; CI B - apresenta os valores do subtteste CI sobreposto ao mapa; CI C - Apresentam os valores predominantes dos valores de CI em cada nodo; TS A - apresenta o mapa de calor para o subtteste TS; TS B - apresenta os valores do subtteste TS sobreposto ao mapa; TS C - Apresentam os valores predominantes dos valores de TS em cada nodo. Percebe-se que a pontuação de CI tende a ser maior à medida em que a série aumenta (comparar as Figuras 4 e 5-CI), enquanto que a distribuição de TS é mais homogênea (comparar as Figuras 4 e 5-TS), com exceção para a área correspondente à maior concentração de estudantes da 1ª série, na parte esquerda do mapa, onde as semelhanças são um pouco maiores. De fato, o coeficiente de correlação de Pearson é moderado para CI (0,67) e desprezível para TS (0,24).



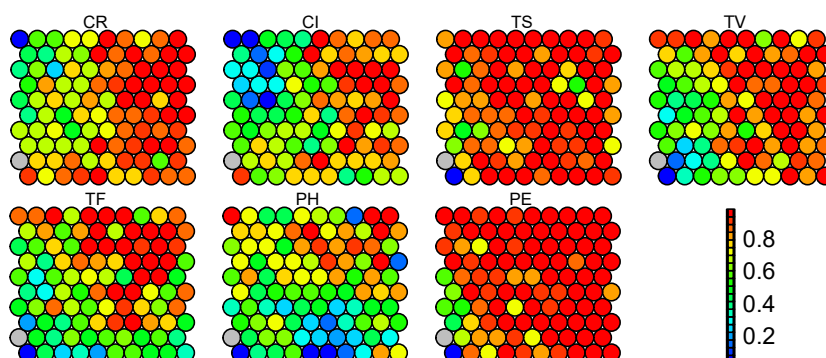
**Figura 5. Mapas dos atributos CI e TS resultantes do treinamento do SOM com os sete atributos do TCLPP. Na coluna A, são representados os mapas de calor de CI e TS (valores médios dos atributos) em cada nodo; na B, os valores de CI e TS sobrepostos ao mapa; na C, os valores predominantes (moda) dos subttestes CI e TS em cada nodo.**

A Figura 6, exibe os mapas de calor para cada um dos sete subttestes do TCLPP resultantes do treinamento do SOM. A variação de cores, do azul para o vermelho, indicam a distribuição topológica dos valores dos subttestes sobre o mapa de características. Através desta visualização, é possível verificar a correlação entre os atributos do TCLPP. Quanto mais parecidos são os mapas, maior é a correlação. Os mapas correspondentes a CR e CI são parecidos, com as maiores pontuações na parte direita do mapa, indicando uma correlação alta (de fato, o coeficiente de correlação de Pearson entre os atributos é considerado alto: 0,73). Já os mapas correspondentes a CI e PE são muito diferentes, pois, os valores mais altos de PE estão distribuídos por todo o mapa, com exceções na



parte esquerda inferior do mapa (o coeficiente de Pearson entre os dois atributos é considerado desprezível: 0,19). Os demais valores de correlação de Pearson entre os atributos do TCLPP são: CR e TS = 0,23; CR e TV = 0,43, CR e TF = 0,30; CR e PH = 0,07; CR e PE = 0,28; CI e TS = 0,17; CI e TV = 0,42; CI e TF = 0,35; CI e PH = 0,14; CI e PE = 0,19; TS e TV = 0,44; TS e TF = 0,33; TS e PE = 0,68; TV e TV e TF = 0,60; TV e PH = 0,29; TV e PE = 0,56; TF e PH = 0,48; TF e PE = 0,45; PH e PE = 0,26.

O conjunto de mapas da Figura 6 revela um *outlier* no nodo inferior esquerdo do mapa (logo abaixo do nodo cinza). Existe um único elemento (estudante) da base de dados associado a este nodo, e ele obteve pontuação máxima (1 - vermelho) para os subtestes CR e CI e nota mínima (0 - azul) para todos os demais, revelando que o estudante assinalou todas as respostas como verdadeiras. Vale ressaltar que, é esperado que os estudantes assinalem os subtestes CI e CR como verdadeiros e todos os demais como falsos.



**Figura 6. Mapas de calor para de cada um dos sete subtestes do TCLPP.**

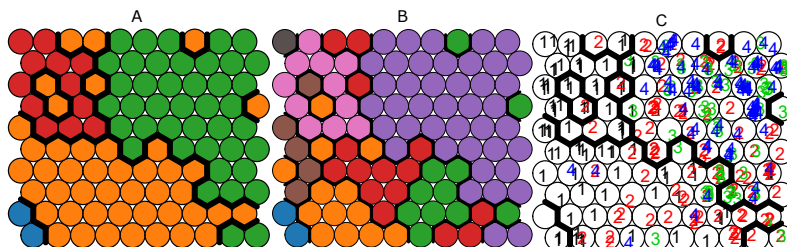
A Figura 7, exibe visões diferentes do mapa de características evidenciando agrupamentos (regiões do mapa, cujos atributos possuem características semelhantes). A Figura 7-A apresenta o mapa dividido em quatro agrupamentos diferentes, enquanto que a Figura 7-B exibe o mesmo mapa com oito agrupamento, sendo que, a maioria deles formada na parte inferior esquerda do mapa, indicando a heterogeneidade dos nodos nesta região do mapa. A quantidade de agrupamentos foi definida com base no critério de menor distância intergrupo e intragrupo, e o recorte sobre o mapa foi obtido utilizando-se técnicas de agrupamento de dados hierárquica [Silva 2015]).

A visualização dos agrupamentos, fornece informações sobre os dados que devem ser investigados. Como a correlação entre o TCLPP e a série escolar é apenas moderada, os agrupamentos não refletem as séries escolares, mas a evolução dos estudantes no processo de ensino aprendizagem que, como pode ser observado, varia bastante entre as séries e mesmo dentro de uma mesma série (ver Figura 7-C). É possível verificar que o TCLPP é melhor em diferenciar os alunos da 1ª e 2ª séries, o que também foi observado na pesquisa de [Dias et al. 2015, pag. 409], utilizando-se técnicas de estatística inferencial.

Outra forma de analisar o comportamento dos dados, é através da análise da concentração dos elementos sobre o mapa. Nas regiões com alta similaridade, os dados tendem a estar concentrados, enquanto que nas regiões com baixa similaridade, dispersos. O mapa da Figura 7-C exibe as informações da série escolar impressas sobre o mapa, permitindo observar as regiões com maior e menor diferenciação com base nos subtestes do TCLPP. Nota-se que na região onde estão concentradas as crianças da 3ª e 4ª séries existe



uma concentração muito grande de elementos, indicando alta similaridade, enquanto que na região onde estão concentradas as crianças da 1ª e 2ª séries, os elementos estão dispersos, indicando dissimilaridade.



**Figura 7. Mapas de características representando respectivamente: A - divisão em quatro agrupamentos; B - divisão em oito agrupamentos; C - quatro agrupamentos sobrepostos com a série escolar.**

## 6. Considerações finais

Este estudo identificou o uso potencial do SOM como ferramenta de análise de exploratória de dados em testes cognitivos como o TCLPP, tendo demonstrado a capacidade da ferramenta em visualizar aspectos obtidos através de análise estatística inferencial.

O SOM mostrou-se particularmente útil como ferramenta de visualização das interações entre os atributos do TCLPP, tanto daqueles que participaram do processo de treinamento do algoritmo, quanto de atributos categóricos, que não participaram do treinamento, como a série escolar, identificando aspectos difíceis de serem inferidos tanto na análise direta dos dados brutos, quanto nos dados tratados estatisticamente.

É possível analisar a interação entre os atributos no mapa de características observando-se o mapa como um todo, olhando-se para regiões específicas do mapa (agrupamentos), ou ainda para nodos específicos, na busca por *outliers*. A comparação entre diferentes representações do mapa de características podem auxiliar na compreensão da relação entre os dados.

Uma possível aplicação do SOM, está em apoiar processo de normalização e validação de testes cognitivos como o TCLPP, para certificar-se de que os diferentes atributos conseguem mensurar as características desejadas, no espaço amostral definido (no caso do TCLPP, meninos e meninas entre a 1ª e 4ª séries e entre 6 e 10 anos).

Como estudos futuros, sugere-se utilizar o SOM para explorar a interação entre diferentes tipos de testes cognitivos relacionados com outras áreas do desenvolvimento cognitivo, além da leitura, que são: raciocínio, matemática, linguagem oral, rotação mental, visomotricidade e funções executivas.

## Referências

- [Baldwin et al. 2014] Baldwin, S. A., Imel, Z. E., Braithwaite, S. R., and Atkins, D. C. (2014). Analyzing multiple outcomes in clinical research using multivariate multilevel models. *Journal of Consulting and Clinical Psychology*, 82(5):920–930.
- [Behrens 1997] Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2):131–160.

- [Capovilla and Seabra 2010] Capovilla, F. and Seabra, A. (2010). *Teste de Competência de: Leitura de Palavras e Pseudopalavras (Manual)*. MEMNON.
- [Dias et al. 2015] Dias, N. M., Montiel, J. M., and Seabra, A. G. (2015). Development and interactions among academic performance, word recognition, listening, and reading comprehension. *Psicologia: Reflexão e Crítica*, 28(2):404–415.
- [Dias and Seabra 2014] Dias, N. M. and Seabra, A. G. (2014). Instrumentos de avaliação de componentes da leitura : investigação de seus parâmetros psicométricos. *Aval. psicol. [online]*, 13(11):235–245.
- [Duncan et al. 2007] Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., and Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6):1428–1446.
- [Good 1983] Good, I. J. (1983). The Philosophy of Exploratory Data Analysis. *Philosophy of Science*, 50(2):283.
- [Harp et al. 1995] Harp, S. A., Samad, T., and Villano, M. (1995). Modeling Student Knowledge with Self-Organizing Feature Maps. *Ieee Transactions on Systems Man and Cybernetics*, 25(5):727–737.
- [Heiman 2011] Heiman, G. W. (2011). *Basic statistics for the behavioral sciences*.
- [Hoaglin 2003] Hoaglin, D. C. (2003). John W. Tukey and Data Analysis. *Statistical Science*, 18(3):311–318.
- [Kaski and Kohonen 1996] Kaski, S. and Kohonen, T. (1996). Exploratory Data Analysis By The Self-Organizing Map: Structures Of Welfare And Poverty In The World. *Neural Networks in Financial Engineering. Proceedings of the Third International Conference on Neural Networks in the Capital Markets*, pages 498–507.
- [Kohonen 2013] Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, 37(2013):52–65.
- [Muñoz and Muruzábal 1998] Muñoz, A. and Muruzábal, J. (1998). Self-organizing maps for outlier detection. *Neurocomputing*, 18(1-3):33–60.
- [Salibian-barrera et al. 2016] Salibian-barrera, M., Aelst, S. V., and Yohai, V. J. (2016). Robust tests for linear regression models based on t-estimates. *Computational Statistics & Data Analysis*, 93:436–455.
- [Schmidt 1996] Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1(2):115–129.
- [Silva 2015] Silva, L. A. (2015). *Mineração de dados: uma abordagem introdutória e ilustrada*. Editora Mackenzie, São Paulo, 1 edition.
- [Wilkinson 1999] Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychological Association*, 54 (8)(8):594–604.