

Aplicação de conceitos de redes complexas para a descoberta de formação de grupos em mapas auto-organizáveis

Thiago Rolemberg¹, Leandro A. Silva²

¹Universidade Presbiteriana Mackenzie
Caixa Postal 01302-907 – São Paulo – SP – Brazil

²Programa de Pós-Graduação em Engenharia Elétrica e Computação

³Faculdade de Computação e Informática, Universidade Presbiteriana Mackenzie

trolemberg@gmail.com, leandroaugusto.silva@mackenzie.br

Abstract. *Neural Networks of the Self-Organizing Maps type, in particular, stands out as one of the clustering algorithms for allowing the analysis of cluster characteristics and the topological relationship between clusters from a lattice of neurons. However, there is still a research gap, which consists of discovering the relationship behind the attributes that lead to the formation of groups. In this sense, this work proposes the use of complex network concepts in order to use the lattice neurons to generate a graph and complement the analysis in the community context, analyzing the formation of groups by measures of centrality.*

Resumo. *Redes Neurais do tipo Mapas Auto-Organizáveis ou SOM (do inglês Self-Organizing Maps), em particular, se destaca como um dos algoritmos de agrupamento por permitir analisar as características de agrupamento e a relação topológica entre grupos a partir de um reticulado de neurônios. Contudo, ainda há uma lacuna de pesquisa, que consiste em descobrir a relação por de trás dos atributos que levam a formação de grupos. Neste sentido, propõe-se neste trabalho o uso de conceitos de redes complexas no sentido de usar os neurônios do reticulado para a geração de um grafo e complementar a análise no contexto de comunidade, analisando a formação de grupos por medidas de centralidade.*

1. Introdução

Agrupamento de dados consiste em uma tarefa de Mineração de Dados cuja aplicação de um algoritmo resulta em objetos de um conjunto organizados em grupos. Isso se dá por meio de medidas de similaridade ou dissimilaridade entre os objetos que são representados por um conjunto de atributos. A tarefa de agrupamento de dados desperta o interesse de diferentes áreas de conhecimento e da indústria pelo fato de permitir descobrir de maneira não supervisionada algum conhecimento do conjunto de dados que auxilie em tomadas de decisões [Z. Wu 1993], [Pei 2011], [Jain 2010]. Exemplos típicos de aplicações vão desde segmentação de imagens [Z. Wu 1993] a descoberta de perfil de usuários em conjunto de dados [Jain 2010].

Na literatura podem ser encontradas algumas formas de taxonomia de algoritmos de agrupamentos, nesta esta pesquisa é destacado as taxonomias: particionais e

hierárquicos [Jain 2010, Saxena et al. 2017]. Os algoritmos particionais separam os objetos a partir de uma parametrização do número de grupos desejado. Por sua vez, os métodos hierárquicos organizam os objetos por similaridade, podendo ter como uma característica a decisão sobre o número de grupos ao usuário ou como parâmetro externo ao algoritmo. Ainda existem os métodos bio inspirados que geralmente tentam descobrir diferentes composições de grupos guiados por alguma função de custo para otimização. Alternativamente aos algoritmos supracitados destaca-se a rede neural de Mapas Auto-Organizáveis. SOM (do inglês *Self-Organizing Maps*) como a rede é conhecida tem uma característica similar de particionamento, agrupando objetos semelhantes em neurônios e, ao mesmo tempo, tem a característica de estruturar os neurônios em um reticulado, organizado de baixa dimensão, em geral, representado em espaço de n-dimensões em 2D, que garante a manutenção da topologia dos dados.

Com o interesse cada vez maior de que os algoritmos de aprendizagem de máquina sejam cada vez menos caixa-preta, conceito popularmente chamado de XAI (do inglês, *eXplainable Artificial Intelligence* [Arrieta et al. 2020]), no contexto de agrupamento um interesse que se torna relevante é saber explicar como se deu uma formação de grupos. Algumas tentativas foram feitas usando para isso conceitos de redes complexas e considerando o reticulado de neurônios como sendo vértices de um grafo. A citar como exemplo a tentativa de geração de um grafo a partir de neurônios com uma abordagem empírica para remover arestas com conexão fraca, gerando particionamentos no reticulado [Silva and Costa 2011]. Outras tentativas ainda considerando SOM e redes complexas foram feitas para criar uma reordenação do reticulado permitindo entender sequências de imagens e vídeos a partir da reorientação do reticulado em um grafo orientado [Kitani et al. 2012, Kitani et al. 2013].

Embora os trabalhos anteriores tenham caminhado para fornecer informações adicionais sobre os grupos a partir de redes complexas, ainda existe a necessidade de conseguir explicar como se dá uma formação de grupos e como os mesmos se relacionam. Nessa lacuna que este artigo se insere. Portanto, o objetivo deste trabalho é apresentar uma abordagem híbrida usando SOM e Redes Complexas em que primeiro se organiza os dados em um reticulado de neurônios e, depois, reorganiza os neurônios em vértices de um grafo usando como informações de centralidade para que se consiga descobrir como a formação de grupos acontece e explicar como os mesmos se relacionam.

Além da introdução, este artigo está organizado em trabalhos correlatos (Seção 2), referencial teórico (Seção 3), método proposto (Seção 4), metodologia experimental (Seção 5), experimentos (Seção 6) e, por fim, conclusões e trabalhos futuros.

2. Trabalhos Correlatos

Durante o desenvolvimento desse artigo podem ser citados alguns trabalhos que usam SOM como um tipo de pré-processamento de dados para aplicação de algoritmos de agrupamento combinados para a descoberta de grupos.

Um dos primeiros trabalhos nesta matéria é um dos mais citados na literatura envolveu realizar a organização dos dados do SOM e conseqüentemente realizar o agrupamento dos vetores de pesos do neurônio usando k-médias (agrupamento particional) e agrupamento aglomerativo (agrupamento hierárquico). No trabalho os autores justificam a abordagem híbrida como forma de eliminar o outlier dos dados e reduzir o número de

objetos a partir do SOM [Vesanto and Alhoniemi 2000].

Outros trabalhos que se assemelham mais da proposta deste trabalho partem da mesma ideia do trabalho de [Vesanto and Alhoniemi 2000]. Contudo usam conceitos de redes complexas ao invés de algoritmos agrupamento de dados como os apresentados em [Saxena et al. 2017].

Como a saída do SOM é um reticulado, pode-se representar cada neurônio como vértice e alguma medida de similaridade entre os neurônios adjacentes para medir a similaridade é uma estratégia em transformar o reticulado em um grafo [Silva and Costa 2011]. A tentativa dos autores foi de conseguir descobrir o número de grupos no conjunto de dados sem a interferência de parâmetros como proposto por [Vesanto and Alhoniemi 2000].

Contudo, ambas as abordagens não se interessavam em descobrir como os grupos se formam. O trabalho mais próximo a isto também partiu do reticulado dos neurônios, transformando em um grafo não orientado de forma a criar uma rede complexa. Por meio de algoritmo de Dijkstra os autores conseguiam percorrer o vértices (neurônio) com a finalidade de entender como uma sequência de imagens podem ser apresentadas [Kitani et al. 2012, Kitani et al. 2013].

A proposta deste trabalho se assemelha as propostas de [Kitani et al. 2012, Kitani et al. 2013]. A diferença que o objetivo aqui é trabalhar com dados numéricos e com a finalidade de descobrir como os grupos se formam e para isto se usa medidas de centralidade, como será apresentado no decorrer deste trabalho.

3. Referencial Teórico

3.1. Mapas Auto-Organizáveis

Mapas Auto-Organizáveis ou simplesmente SOM (do inglês, *Self Organizing Maps*) consiste de um tipo de Redes Neurais Artificiais com arquitetura de duas camadas, sendo a primeira para recebimento dos dados de entrada e a segunda formada por um reticulado com neurônios. Na maioria dos casos esse reticulado é bidimensional (2-D). A estrutura do reticulado 2-D pode ser ainda hexagonal ou regular [Kohonen 2013].

A camada de entrada espera receber um objeto \mathbf{x}_i de um conjunto de dados $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N\}$. Cada objeto é definido por um conjunto de M atributos, $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{iM}\}$.

Na camada de saída, o reticulado é parametrizado com U unidades, sendo cada neurônio u definido como um vetor de pesos com a mesma dimensão do objeto de entrada, ou seja, $\mathbf{w}_u = \{w_1, w_2, \dots, w_M\}$. Estes pesos são inicializados de forma aleatória e ajustados durante o aprendizado.

O treinamento do algoritmo SOM é realizado de forma iterativa. Inicialmente, o valor de $t = 0$, os vetores de peso são inicializados aleatoriamente, de preferência a partir do domínio de vetores de entrada [Kohonen 2001]. Em cada etapa do treinamento t , um padrão de entrada $\mathbf{x}_i(t)$ é escolhido aleatoriamente do conjunto de treinamento \mathbf{X} . Cada objeto comparado com os vetores de peso por meio de uma medida de distância, geralmente Euclidiana. A distância entre $\mathbf{x}_i(t)$ e todos os vetores de peso $(w)_u$ são calculadas. O neurônio vencedor é o protótipo mais próximo de $\mathbf{x}_i(t)$ sendo este o *Best Match Unit*

(BMU). O vetor de peso BMU é atualizado, assim como o vetor de pesos dos neurônios dos seus vizinhos, mas com a intensidade menor. Para detalhe completo do algoritmo recomenda-se a leitura desse artigo [Kohonen 2013].

Cabe por fim ressaltar que após o treinamento, os neurônios representam um sub-grupo do conjunto de dados que tem semelhanças em aspectos de atributos e, por sua vez, neurônios vizinhos representam conjunto de dados que mantém algumas características em comum. Esta manutenção topologia dos dados é uma das principais razões para uso do SOM em problemas de agrupamento de dados. Contudo, é importante mencionar que se trata de um reticulado fixo em que situações onde os dados são diferentes em termos de atributos, ou seja, grupos distintos, ele se mantém próximos no espaço do reticulado. E isso, por sua vez, é uma característica do SOM que dificulta o entendimento da formação de grupos e quais os atributos similares.

3.2. Redes Complexas

O estudo de redes complexas trabalha com conceitos de estatística, sistemas dinâmicos e teoria dos grafos [Monteiro 2014]. Muitas áreas têm se interessado nesse estudo como física, matemática, biologia e sociologia, devido a suas aplicações sobre uma grande variedade de problemas, os quais incluem redes sociais, redes biológicas, Internet e World Wide Web e redes de energia elétrica [Newan 2005, Barabasi 2002, Strogatz 2001]. As características topológicas dessas redes não são triviais, nem completamente regulares nem completamente aleatórias, por isso são denominadas de redes complexas. Medidas são utilizadas para caracterizar a estrutura dessas redes. A utilização das medidas são, frequentemente, usadas para analisar as propriedades estatísticas que descrevem a estrutura e o comportamento de sistemas em rede, enquanto a criação de modelos de rede está normalmente relacionada ao entendimento do significado dessas propriedades.

A utilização de redes complexas para análise de dados, geralmente é feita por grafos (modelados, representados) onde cada vértice representa um nó da rede (pessoa, animal, empresa, objeto, elemento) e cada aresta representa qualquer relação arbitrária entre entidades [Grando 2015].

A definição de um grafo é dada por $G = (V, A)$, onde V define o conjunto de nós ou vértices e A é o conjunto de arestas, conexões ou ligações [Monteiro 2014, Boaventura Netto 2001].

Num grafo não direcionado ou não dirigido, a aresta que conecta dois vértices quaisquer i e j também conecta j a i ; portanto, o par (i, j) é não ordenado. Num grafo direcionado ou dirigido, a presença de aresta que parte de i e chega a j não implica a existência de aresta que parte de j e chega a i ; ou seja, o par (i, j) é ordenado.

3.2.1. Medidas Centralidades de Redes Complexas

A análise de uma rede envolve entender aspectos importantes como quais são os vértices mais importantes ou centrais. As medidas de centralidade são uma forma de quantificar essa importância. As medidas de rede são utilizadas para inferir um grafo e destacar a relevância de uma aresta, de um vértice ou definir as características e o tipo de grafo [Monteiro 2014, Grassi 2009].

A centralidade de grau (em inglês *degree centrality*) é considerada a mais simples de todas as medidas, ela avalia a importância de um nó no grafo, analisando a quantidade de nós a que ele é ligado, ou seja, quanto maior o número de nós ligados a este, maior a importância dele e, portanto, maior o valor atribuído para este [Borgatti 1999]. A centralidade k o i ésimo vértice é calculada como [Grassi 2009]:

$$k_i = \sum_{j=1}^n A_{ij} \quad (1)$$

sendo que i e j são elementos da matriz adjacente A da rede e n o número de vértices na rede.

A medida de intermediação (em inglês *betweenness centrality*) consiste em avaliar a importância de um nó na transmissão de mensagens ou eventos entre os demais, ou de maneira equivalente, como ele se encontra no caminho entre os outros vértices da rede se quiserem trocar informações [Newan 2005, Grandó 2015]. A intermediação de um nó i é dada por:

$$g(v) = \sum_{i \neq v \neq j} \frac{g_{ij}(v)}{g_{ij}} \quad (2)$$

onde $g_{ij}(v)$ é o número total de menores caminhos do vértice i para o j e g_{ij} é o número de menores caminhos que passam por v .

A centralidade de proximidade (em inglês, *closeness centrality*) proposta por [Grassi 2009] tem como objetivo avaliar o quanto um determinado vértice está distante dos demais. Assim, os vértices que possuem uma menor distância média comparados com os demais, receberão um valor alto para a centralidade [Grassi 2009].

A centralidade de proximidade é calculada como sendo [Grassi 2009]:

$$C_i = \frac{1}{l_i} \quad (3)$$

$$l_i = \frac{1}{n-1} \sum_{j(\neq i)} d_{ij} \quad (4)$$

sendo que n representa o número total de vértices na rede; d_{ij} é o comprimento do menor caminho entre os vértices i e j ; l_i representa a média do comprimento das menores distâncias entre i e todos os outros vértices da rede.

4. Método proposto de combinação do SOM com Redes Complexas

O método proposto nesta pesquisa está ilustrado na Figura 1. Na figura mais à esquerda o resultado do SOM está ilustrado com um reticulado em 2D, em que se procurou ilustrar como a manutenção topológica dos dados acontece, permitindo que casos parecidos permaneçam entre neurônios próximos. É possível observar que em cada quadrante, existem algumas barras de cores que representam como ficou o agrupamento dos dados após a execução do SOM.

Como se pode notar, o reticulado é uma grade fixa em que para casos em que os neurônios vizinhos representam objetos semelhantes, a interpretação e formação de grupos é natural de fazer. Contudo, para situações em que começam a aparecer dissimilaridade entre os objetos isso pode levar a falsas interpretações pelo fato dos neurônios serem adjacentes no reticulado.

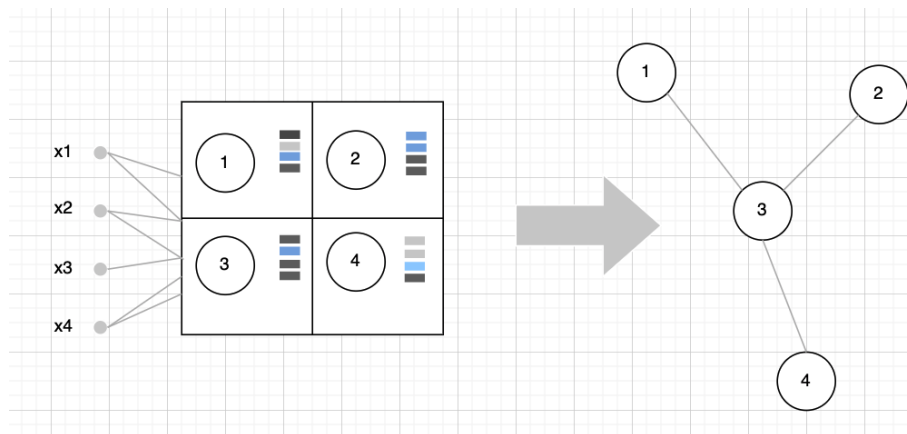


Figura 1. Representação do processo de transformação do SOM para o grafo.

A proposta deste trabalho é que após treinamento do SOM, o mapa seja transformado em um grafo não direcionado, onde cada neurônio será representado por um vértice e as arestas representada da seguinte maneira:

$$a_{i,j} = \frac{1}{dist(\mathbf{w}_i, \mathbf{w}_j)} \quad (5)$$

sendo $a_{i,j}$ a aresta entre os vértices i e j , dois neurônios do mapa, $dist()$ a distância Euclidiana. A proposta de medir pelo inverso da distância é para que menores distâncias sejam transformadas em maiores pesos. Além da proposta na forma de calcular a aresta, este trabalho ainda contempla a eliminação de arestas com valores pequenos. De forma empírica arestas com quartil menor a 25% serão removidas.

5. Materiais e Métodos

A Figura 2 apresenta o processo completo de como se dá o experimento usando a proposta apresentada neste trabalho. Todo o processo inicia com a base de dados, depois o mapa SOM é parametrizado adotando aqui a configuração quadrada de dimensão 3×3 , para trabalhos futuros pretende-se explorar outras configurações com um numero maior de dimensões. Uma vez o mapa treinado calcula-se as arestas e calculam-se as medidas de centralidade. A interpretação da formação de grupos é apoiada por uma metodologia que permite visualizar por ferramentas gráficas do SOM e de grafos.

Para a realização deste experimento foi utilizado a base de dados Animal [Dua and Graff 2017]. A base é composta por características como definidas na Tabela 1 que definem um determinado animal em: 1-Mamíferos, 2-Aves, 3-Répteis, 4-Peixes, 5-Anfíbios, 6-Insetos e 7-Moluscos e Crustáceos. A tabela 1 representa a distribuição das classes da base de dados animais e sua legenda. Importante ressaltar que a classe é o atributo type na base.

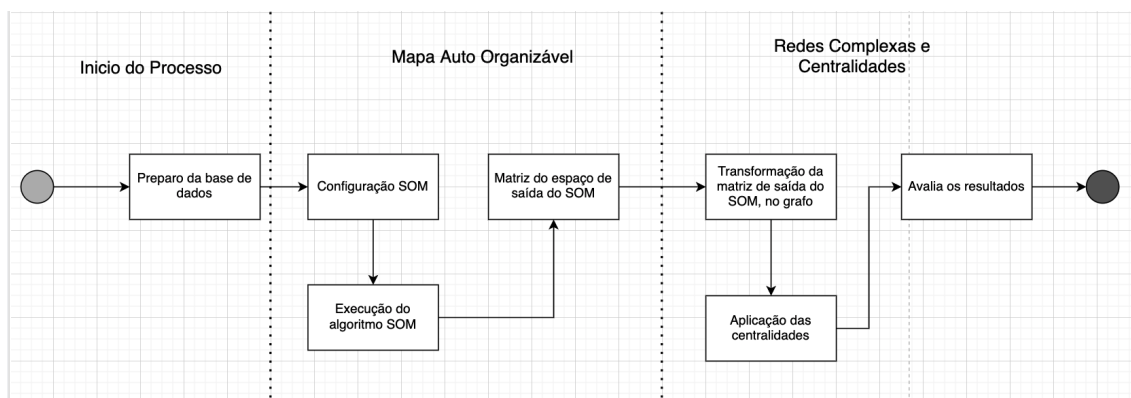


Figura 2. Representação do processo de transformação do SOM para o grafo.

Classe	1	2	3	4	5	6	7
Distribuição	41	20	13	13	4	8	10

Tabela 1. Classe dos animais e Distribuição.

Todo o código desenvolvido usará a linguagem de programação R, com as seguintes bibliotecas. *SOMbrero* que é uma biblioteca da linguagem R para treinamento da rede SOM [Olteanu and Villa-Vialaneix 2015, Olteanu et al. 2012, Mariette et al. 2017]. E a biblioteca *Igraph* que contém métodos para geração de grafos e execução dos cálculos de centralidades [Csardi and Nepusz 2006].

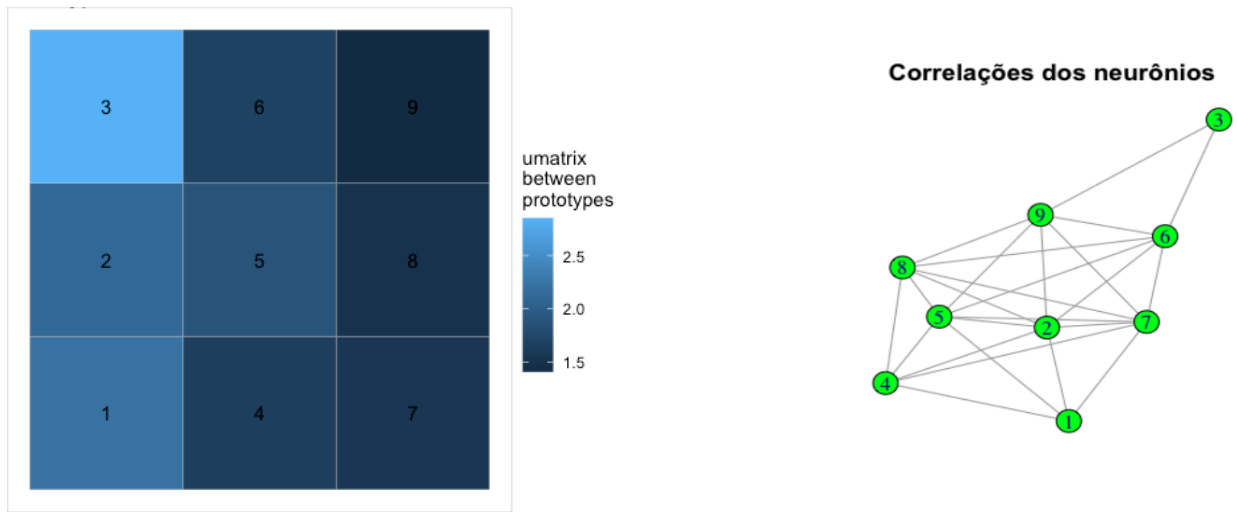
6. Experimento e Resultado

A Figura 3 sintetiza os principais resultados do experimento deste trabalho. A Figura 3a traz três informações importantes para o entendimento do treinamento do SOM. Nessa se pode observar a quantidade de neurônio e suas identificações (de 1 a 9) e a relação topologia dos neurônios adjacentes a partir da medida de distância (Euclidiana) de um neurônio e seus adjacentes. Assim, valores menores e em tons azuis escuros significam neurônios próximos no espaço de vetores de pesos. Cores com intensidades mais claras significam neurônios mais distantes entre si. Assim, nota-se que o neurônio 3 é o mais distante de seus vizinhos 2, 5 e 6. Por sua vez, o neurônio 9 é o que tem o menor valor de distância e, por sua vez, está mais próximo dos seus vizinhos 5, 6 e 8.

Embora os resultados da Figura 3a são valiosos na descoberta de quantidade de grupos, uma vez que neurônios próximos ou distintos possam ser trabalhados no sentido de descobrir automaticamente o número de grupos, esse não permite analisar a relação entre os grupos e a sua formação.

Por isto, há um mapeamento dos neurônios em um grafo, conforme Figura 3b, em que os neurônios viram vértices e a aresta é definida pelo inverso da distância como indicado na Eq. 5. Ou seja, a Figura 3c ilustra uma matriz simétrica relacionando a aresta de um vértice com outros.

A partir do grafo (Figura 3b) se pode fazer algumas associações com a matriz-U como é o caso do neurônio/vértice 3 como sendo o mais afastado. Isso para o SOM indica a maior distância e para o grafo o menor peso. Contudo, a vantagem que se destaca na



(a) U-Matrix cuja intensidade de cores indica a distância média entre vetores de pesos do neurônios adjacentes. (b) Grafo tendo com vértices os neurônios do mapa SOM, numeração vem da figura a)

Vértices	1	2	3	4	5	6	7	8	9
1	0	0,409	0	0,733	0,353	0	0,369	0	0
2	0,409	0	0	0,642	0,928	0,464	0,604	0	0,41
3	0	0	0	0	0	0,551	0	0	0,36
4	0,733	0,642	0	0	0,556	0	0,724	0,463	0
5	0,353	0,928	0	0,556	0	0,536	0,596	1,406	0,58
6	0	0,464	0,551	0	0,536	0	0,347	0,636	0,97
7	0,369	0,604	0	0,724	0,596	0,347	0	0,629	0,33
8	0	0	0	0,463	1,406	0,636	0,629	0	0,69
9	0	0,412	0,363	0	0,576	0,968	0,33	0,685	0

(c) matriz de arestas cujo cálculo está definido na Eq. 5. Valores pequenos indicam vértices sem conexão ou com baixa ligação e valores altos representam conexões fortes.

Figura 3. Resultado experimental com a base animais apresentado pelo reticulado da rede SOM junto com as cores da U-Matrix, a geração do grafo e, por fim, o mapa de cores (Heatmap) com as arestas do grafo.

proposta deste trabalho é justamente a relação dos vértices com os demais, fazendo uma metáfora como sendo uma comunidade de grupos e como são as interações.

Neste sentido, justifica-se o uso das medidas de centralidade apresentadas na Tabela 2. Estas medidas ajudam a explicar o relacionamento da formação dos grupos com será apresentado a seguir. Antes disso, cabe apenas apresentar que cada neurônio e vértice tem uma densidade de objetos agrupados. A distribuição desta densidade (objetos por neurônio/vértice) está apresentada na Tabela 4.

Analisando cada um dos vértices, mais especificamente os objetos que foram mapeados em cada neurônio da rede, conforme Tabela (4) observa-se que o vértice 1 é formado por animais de classe peixe (13), moluscos e crustáceos (2), mamíferos (3) e répteis (3). Analisando as conexões e arestas (Figura 3c) é possível observar que as conexões acontecem com os vértices:

- 2: não tem formação de grupos
- 4: grupo formado predominantemente com moluscos e crustáceos;

Vértice	1	2	3	4	5	6	7	8	9
Grau	7	5	7	7	6	6	6	4	2
Betweenness	2	0.25	2	2	3	0.75	3	0	0
Proximidade	0.111	0.083	0.111	0.111	0.100	0.100	0.100	0.077	0.062

Tabela 2. Medidas de Centralidades.

neurônio	1	2	3	4	5	6	7	8	9
número de objetos	21	0	14	2	7	8	20	0	29
Distribuição dos objetos por classes	4(3), 7(2), 1(13), 3(3)	-	6(8), 7(6)	7(2)	1(7)	5(4), 1(2), 3(2)	2(20)	-	9(29)

Tabela 3. Distribuição de objetos e classes por neurônios.

	1	2	3	4	5	6	7	8	9
pelo	0.05	0	0.29	0	1	0.25	0	0	1
penas	0	0	0	0	0	0	1	0	0
ovos	0.81	0	0.03	1	0	0.88	1	0	0
leite	0.14	0	0	1	1	0.25	0	0	1
voador	0	0	0.43	0	0.29	0	0.8	0	0
aquatico	0.86	0	0.36	0	0.14	0.62	0.3	0	0.03
predador	0.81	0	0.5	0	0.29	0.62	0.45	0	0.55
dentado	0.9	0	0	0	1	0.75	0	0	1
respira	0.24	0	0.64	1	1	1	1	0	1
venenoso	0.19	0	0.21	0	0	0.12	0	0	0
barbatana	0.76	0	0	0	0.14	0	0	0	0
pernas	0	0	2,4	0	2	4	2	0	4
rabo	0.86	0	0.07	0	0.71	0.5	1	0	0.93
domestico	0.05	0	0.07	0	0.14	0.12	0.15	0	0.21
catsize	0.33	0	0.07	0	0.57	0.25	0.3	0	0.83

Tabela 4. Vetores de pesos após treinamento da rede SOM.

- 5 grupo formado predominantemente com mamíferos;
- 7 grupo formado predominantemente com anfíbios.

Por sua vez, a Tabela 4 possibilita entender as características de cada neurônio por meio dos valores após treinamento do vetor de pesos. Por exemplo, analisando as características entre neurônios 1 e 5 nota-se que tem em comum pelo, leite, dente, barbatana. Portanto, características de mamíferos, justificando assim a relação com aresta de 0.353. Por sua vez, a relação de 1 e 7, com aresta de 0.369 a característica em comum é o ovo.

Em relação às centralidades, estudadas nesta pesquisa. Para este modelo os vértices 1, 3, e 4 apresentaram maiores valores de grau e proximidade, vide Tabela 2. Pegando novamente o exemplo do vértice 1, formado por peixe, moluscos e crustáceos, mamíferos e répteis, pode-se concluir que esses animais são os que mais se destacam como importantes para rede por serem os que mais se aproximam dos demais animais. Já para a centralidade de *betweenness* os vértices 5 mamíferos e vértice aves, logo pode se concluir que esses animais são considerados responsáveis por estarem mais nos caminhos da formação dos demais grupos do grafo.

7. Conclusão e Trabalhos Futuros

A necessidade de estudos que utilizam métodos e agrupamento de dados tem recebido interesse da comunidade científica no aspecto que a tomada de decisão é de fácil interpretação pelo ser humano. Entender a formação de grupos, por outro lado, ainda é um desafio aos estudos relacionados à descoberta de grupos.

Neste sentido, a transformação do espaço de saída do SOM para modelagem por redes complexas permite estudar as relações que cada vértice tem em relação aos demais e a utilização de medidas de centralidades que avaliam a importância dos vértices e as ligações da rede. Assim, torna-se uma técnica complementar para a descoberta da formação de grupos.

Neste artigo foi proposto o uso de redes complexas para a descoberta de formação de grupos em um SOM, analisando as relações dos vértices e as centralidades de rede. Para o estudo foi usada uma base de dados conhecida como Animal para permitir uma explicação simples do potencial da proposta aqui apresentada, possibilitando ilustrar as relações de grupos e a importância das características da formação de grupos.

Além da forma como as técnicas foram combinadas, outra contribuição do trabalho foi na maneira como é feita a transformação do espaço de saída do SOM em um modelo de grafo. Para tanto, o peso entre vértices foi calculado como sendo o inverso da distância entre os pesos dos neurônios do espaço de saída do SOM. Outra contribuição foi eliminar arestas com neurônios que não fazem vizinhança com todos os neurônios do reticulado. Para isso usou-se a heurística por percentil 25%, para trabalhos futuros pretende-se usar outras abordagens, mantendo assim um modelo de rede mais próximo do que representa o espaço de saída do SOM.

O resultado relatado no experimento mostrou que é possível transformar o espaço de saída do SOM em um modelo de redes complexas e estudar tanto as conexões do grafo como as centralidades, para gerar o entendimento da formação dos grupos.

Como trabalho futuro pretende-se aplicar a proposta em uma base de dados real com mais objetos e atributos de forma a validar a proposta. Em complemento, pretende-se

ainda analisar outras formas de remoção de arestas por quartéis e, analisar o impacto em outras dimensões da rede SOM. Pretende-se por fim sistematizar o processo para que o resultado possa ser feito como uma única saída, sem a necessidade de interpretações pelo usuário.

Referências

- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115.
- Barabasi, A.-L.; Jeong, H. N. Z. R. E. S. A. V. T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, v. 311:p.590–614.
- Boaventura Netto, P. O. (2001). *Grafos: Teoria, Modelos, Algoritmos.*, volume 2. Editora Blucher.
- Borgatti, S. P., E. M. G. (1999). The centrality of groups and classes. *Journal of Mathematical Sociology*, 23:181–201.
- Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems:1695.
- Dua, D. and Graff, C. (2017). UCI machine learning repository.
- Grando, F. (2015). On the analysis of centrality measures for complex and social networks. <http://hdl.handle.net/10183/122516>.
- Grassi, R., S. S. T. A. (2009). Centrality in organizational networks. *International Journal of Intelligent Systems*, 25:253–265.
- Jain (2010). Data clustering: 50 years beyond K-Means. *Pattern Recognition Letters*, 31(8):651–666.
- Kitani, E. C., Del-Moral-Hernandez, E., and Silva, L. A. (2012). Somm–self-organized manifold mapping. In *International Conference on Artificial Neural Networks*, pages 355–362. Springer.
- Kitani, E. C., Del-Moral-Hernandez, E., and Silva, L. A. (2013). Learning embedded data structure with self-organizing maps. In *Advances in Self-Organizing Maps*, pages 225–234. Springer.
- Kohonen, T. (2001). Self-Organizing Maps. Third extended. *Springer, Heidelberg*.
- Kohonen, T. (2013). Essentials of the self-organizing map. *Neural networks*, 37:52–65.
- Mariette, J., Rossi, F., Olteanu, M., and Villa-Vialaneix, N. (2017). Accelerating stochastic kernel som. In M., V., editor, *Proceedings of XXVth European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2017)*, pages 269–274. i6doc.
- Monteiro, L. H. A. (2014). *Sistemas Dinâmicos Complexos.*, volume 2. Editora Livraria da Física.

- Newan, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, 27:39–54.
- Olteanu, M. and Villa-Vialaneix, N. (2015). On-line relational and multiple relational som. *Neurocomputing*, 147:15–30.
- Olteanu, M., Villa-Vialaneix, N., and Cottrell, M. (2012). On-line relational som for dissimilarity data. In P., E., J., P., P., Z., and G., B., editors, *Advances in Self-Organizing Maps (Proceedings of WSOM 2012, Santiago, Chili, 12-14 decembre 2012)*, volume 198 of *Advances in Intelligent Systems and Computing series*, pages 13–22, Berlin/Heidelberg. Springer Verlag.
- Pei, J. H. M. K. J. (2011). *Data Mining: Concepts and Techniques*. Elsevier, Waltham, MA, USA.
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., and Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, 267:664–681.
- Silva, L. A. and Costa, J. A. F. (2011). A graph partitioning approach to som clustering. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 152–159. Springer.
- Strogatz, S. H. (2001). Exploring complex networks. *nature*. *Nature Publishing Group*, v. 410:p.268–276.
- Vesanto, J. and Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on neural networks*, 11(3):586–600.
- Z. Wu, R. L. (1993). An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1101–1113.