  
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA  
DOUTORADO EM MÉTODOS DE APOIO À DECISÃO

**TRABALHO 3: MAPA DE KOHONEN**

**Gabriel Baruque**

**2012325**

Rio de Janeiro

2020

GABRIEL BARUQUE

**TRABALHO 3: MAPA DE KOHONEN**

Trabalho apresentado à disciplina ELE2394 – Redes Neurais I, do programa de Pós Graduação em Métodos de Apoio à Decisão da Coordenação de Engenharia Elétrica da PUC-Rio como requisito parcial de avaliação.

Professora: Marley Vellasco

Rio de Janeiro

2020

# Introdução

Este trabalho tem como objetivo utilizar uma base de dados clínica sobre problemas na tireoide e agrupar seus dados da melhor maneira possível utilizando mapas de kohonen, a fim de determinar o perfil de cada grupo obtido. Está dividido em duas etapas:

* Etapa 1 – Configuração: diversas configurações com relação à topologia, normalização dos dados de entrada e tempo de treinamento das fases de ordenação e ajuste fino, serão testadas, e por fim, uma delas selecionada como a melhor.
* Etapa 2 – Análises: à luz da configuração selecionada na etapa anterior, os resultados obtidos serão analisados e os grupos obtidos serão caracterizados.

Parâmetros fixos escolhidos:

* Taxa de aprendizado (Ordering): 0,97
* Taxa de aprendizado (Fine Tuning): 0,01
* Alpha Type: linear

Sobre os parâmetros fixos, a taxa de aprendizado da fase de ordenação foi escolhida com esse valor pois é necessário que esteja próximo de 1 (um valor alto) e decresça monotonicamente. Já, na fase de “fine tuning”, é necessário que seja um valor pequeno (menor que 0.02), nesse caso, próximo de 0.

# Resultados

**1º teste:**

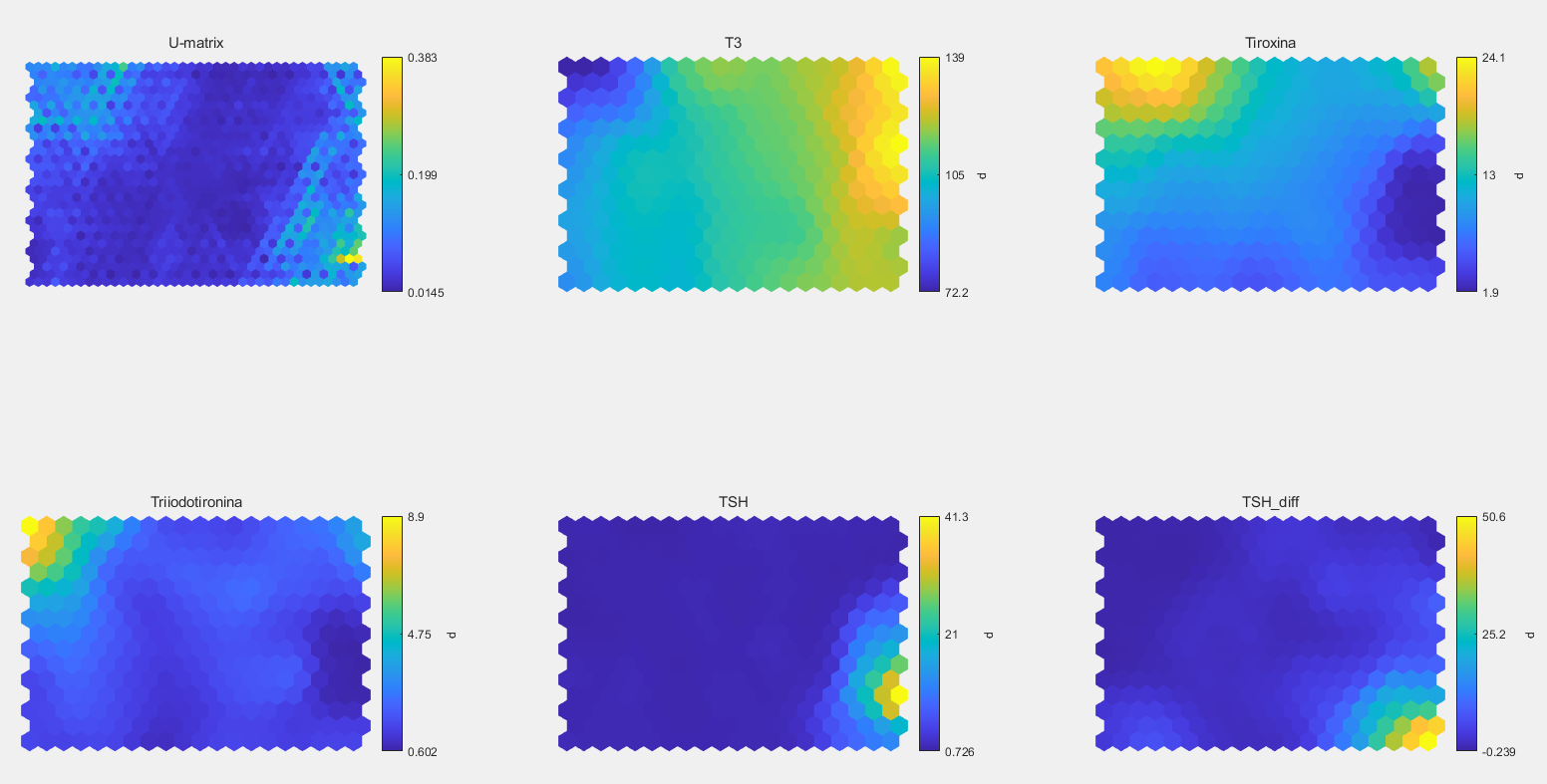
Parâmetros:

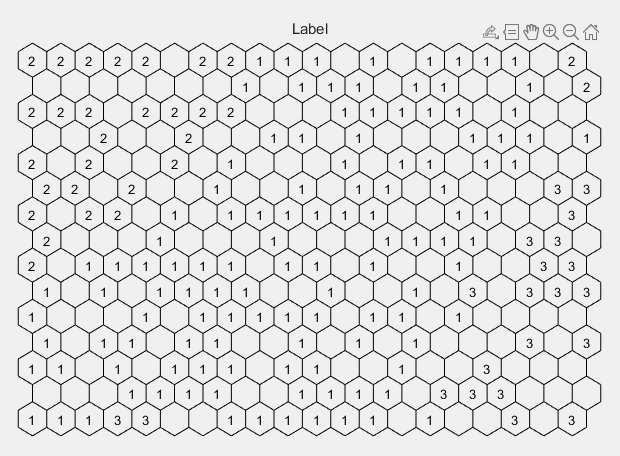
* Topologia do mapa: 15x20
* Raio: [13 1]
* Épocas na Ordenação: 200
* Épocas no Fine Tuning: 150000
* Normalização: ‘range’

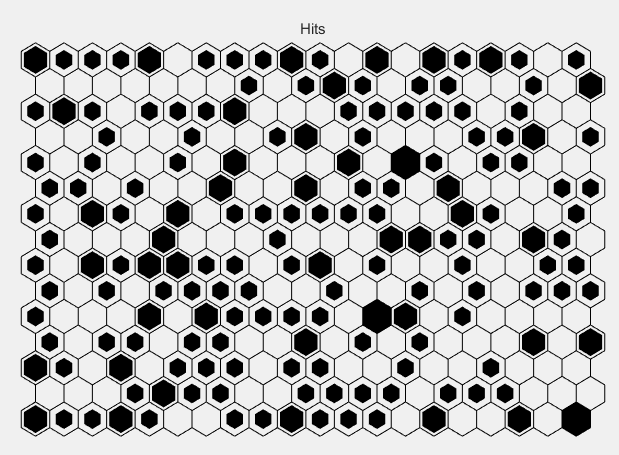
Resultados:

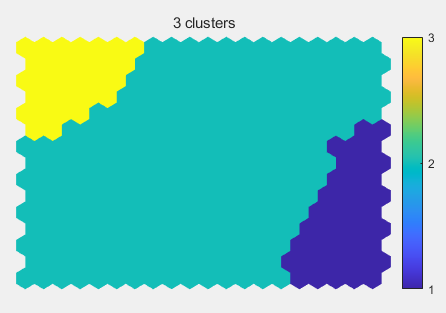
Final quantization error: 0.048

Final topographic error: 0.005

Gráficos: 







No primeiro experimento, foi testado um mapa com uma topologia de 15 x 20, e raio de 13, diminuindo conforme a fase de ordenação passava, chegando em 1. O número de épocas de ordenação foi de 200, e de fine tuning, de 500 vezes o número de processadores, como mencionado em aula. Já foi possível observar um mapa bem dividido, mesmo com uma quantidade de épocas relativamente pequena na fase de ordenação. Grupos em pontas distintas da matriz U são formados. Observando-se distribuições nos mapas relacionados a cada atributo, também fica claro a divisão desses grupos. Porém, ao se observar os processadores ativados no mapa de “Hits”, nota-se que as divisões entre os grupos não são tão bem definidas, sendo necessário mais testes.

O gráfico de “clustering” dos processadores corrobora a criação de 3 grupos distintos, conforme visto na matriz U.

**2º teste:**

Parâmetros:

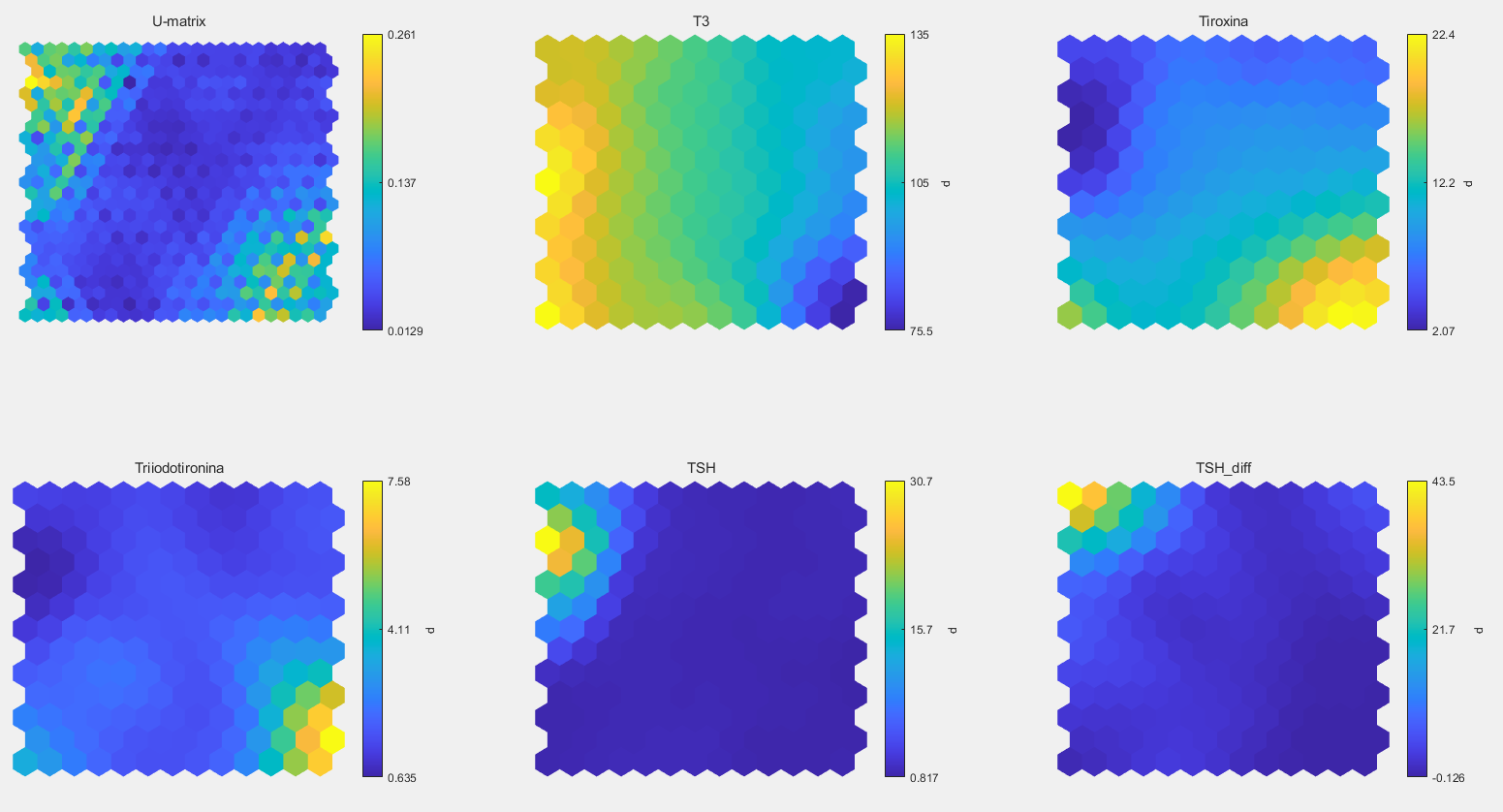
* Topologia do mapa: 13x13
* Raio: [8 1]
* Épocas na Ordenação: 500
* Épocas no Fine Tuning: 84000
* Normalização: ‘range’

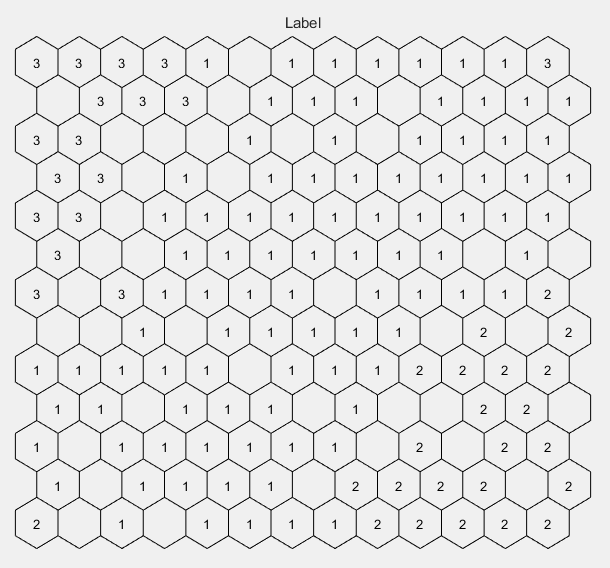
Resultados:

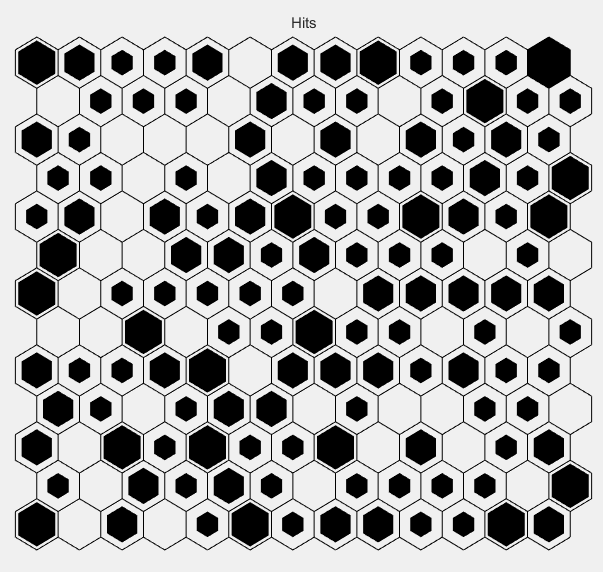
Final quantization error: 0.065

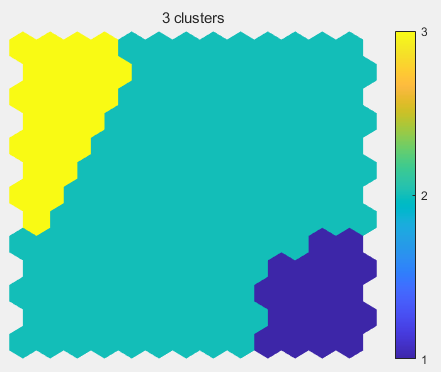
Final topographic error: 0.028

Gráficos:









Neste Segundo teste, um mapa de dimensões menores foi utilizado, para tentar evitar os diversos espaços sem processadores ativados não só entre os grupos, mas também dentro deles. Nesse caso, o grupo 3 obteve uma divisão muito satisfatória, por ser bem definida. Em contrapartida, o grupo 2, assim como em outros testes demonstrou uma dificuldade maior para sua discriminação em relação ao grupo 1.

Embora as métricas de erros tenham sido piores que no primeiro teste, visualmente os mapas gerados parecem ser mais bem definidos, e por esse motivo esta configuração foi selecionada como a melhor do conjunto de testes.

Além disso, é importante notar que a quantidade de épocas para a fase de ordenação foi de 500, um valor abaixo do esperado (em torno de 1000 épocas).

**3º teste:**

Parâmetros:

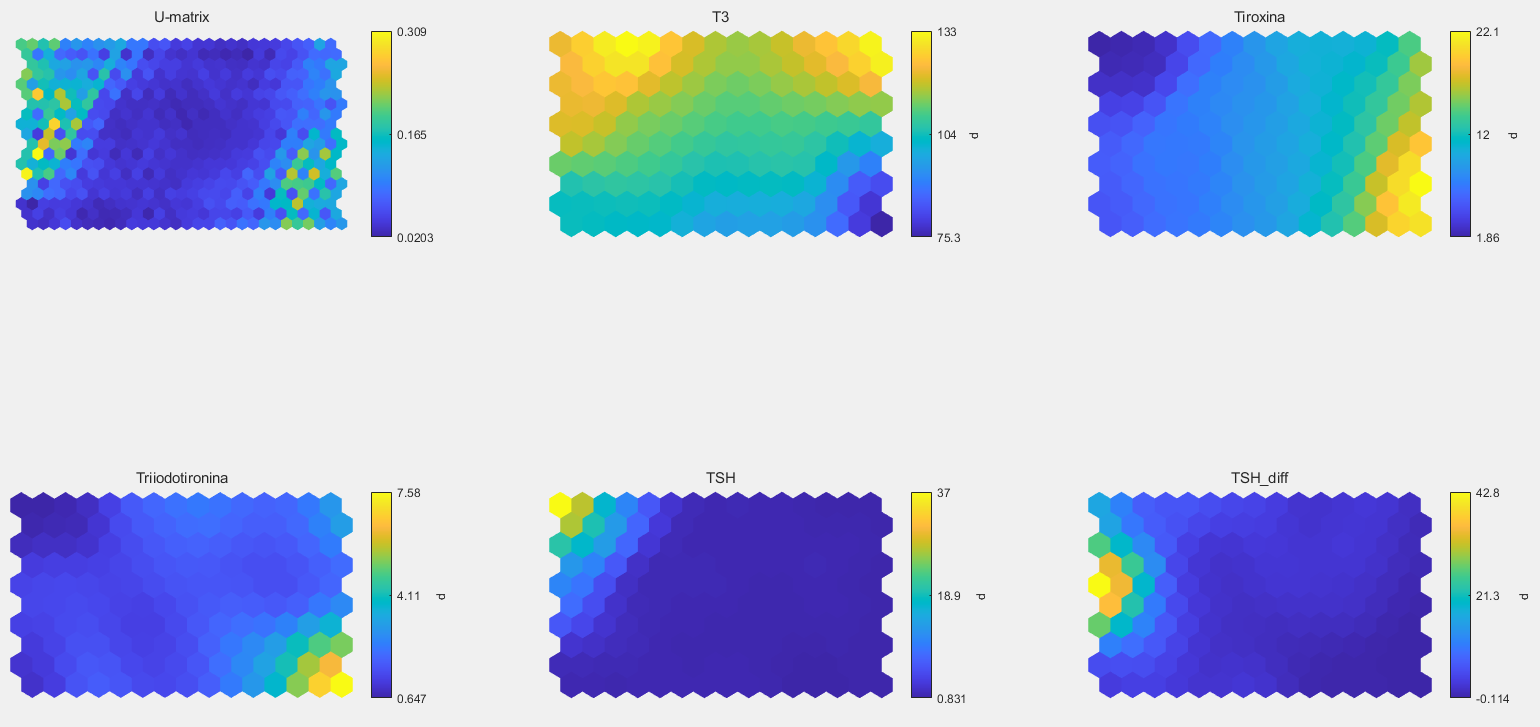
* Topologia do mapa: 10x15
* Raio: [7 1]
* Épocas na Ordenação: 800
* Épocas no Fine Tuning: 75000
* Normalização: ‘range’

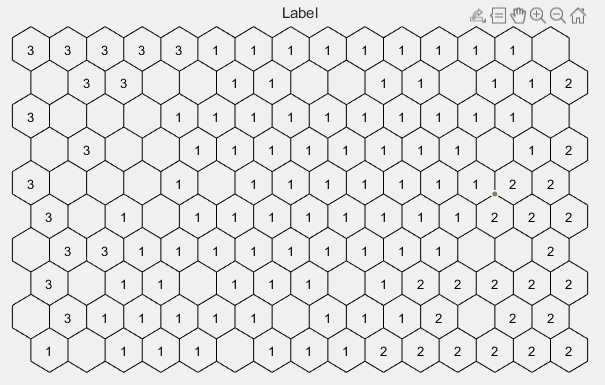
Resultados:

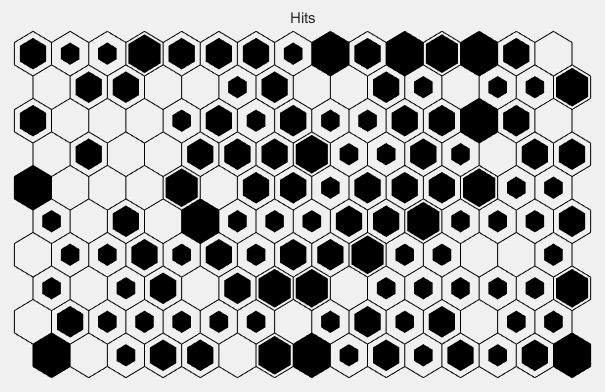
Final quantization error: 0.066

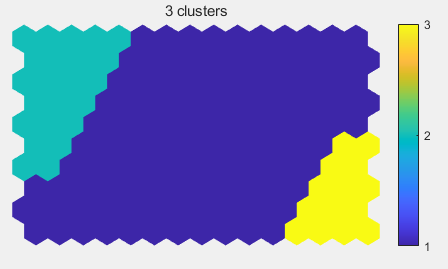
Final topographic error: 0.033

Gráficos:









Para testar mais configurações que pudessem melhorar o resultado dos mapas, outros testes foram realizados. Nesse caso, um mapa de 10x15.

Os resultados das métricas de erros foram ligeiramente piores que o teste anterior, e também, visualmente os mapas estavam menos bem definidos, onde o grupo 3 estava mais mesclado ao 1, assim como o grupo 2.

**4º teste:**

Parâmetros:

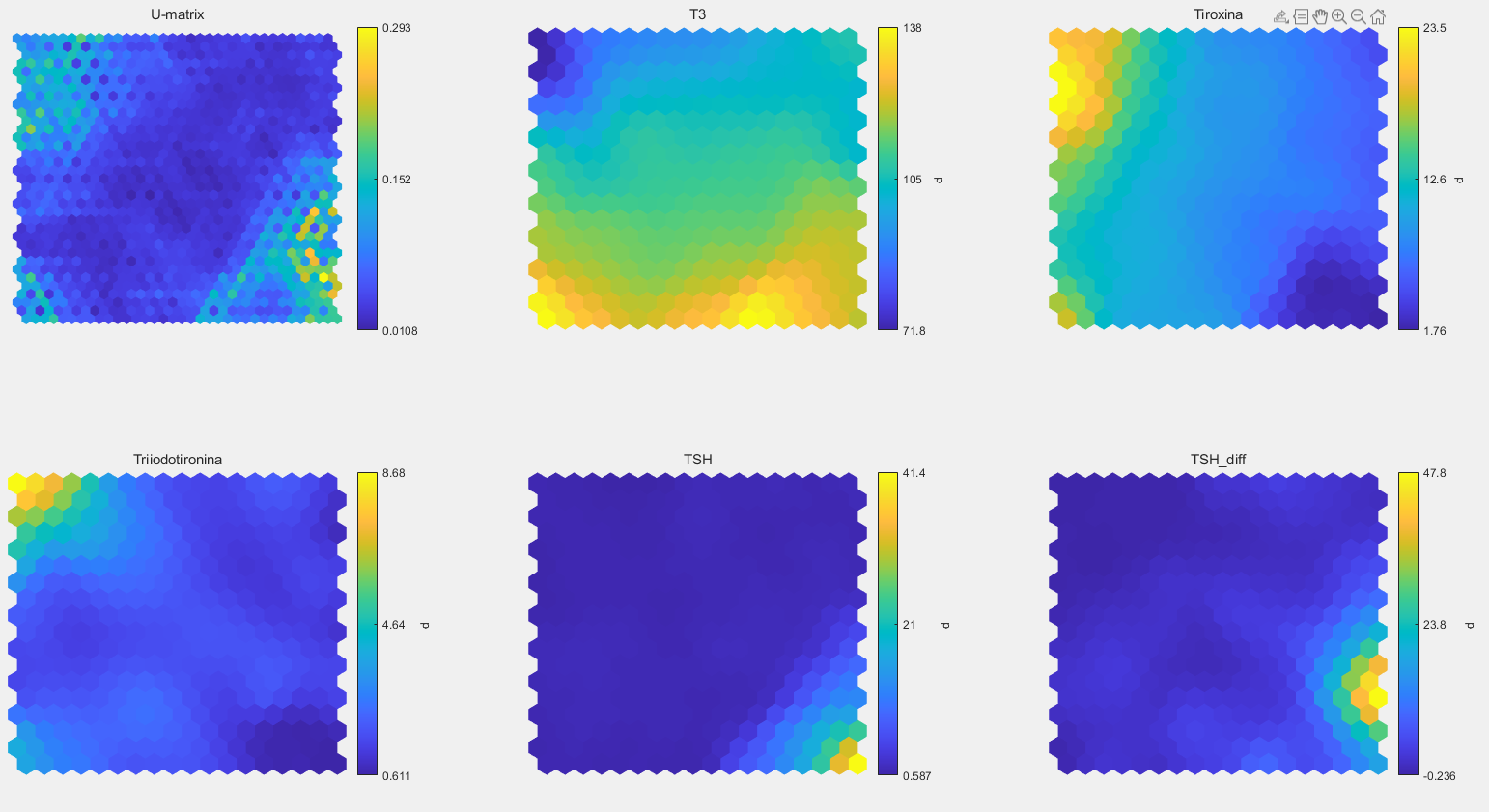
* Topologia do mapa: 18x18
* Raio: [13 1]
* Épocas na Ordenação: 1100
* Épocas no Fine Tuning: 162000
* Normalização: ‘range’

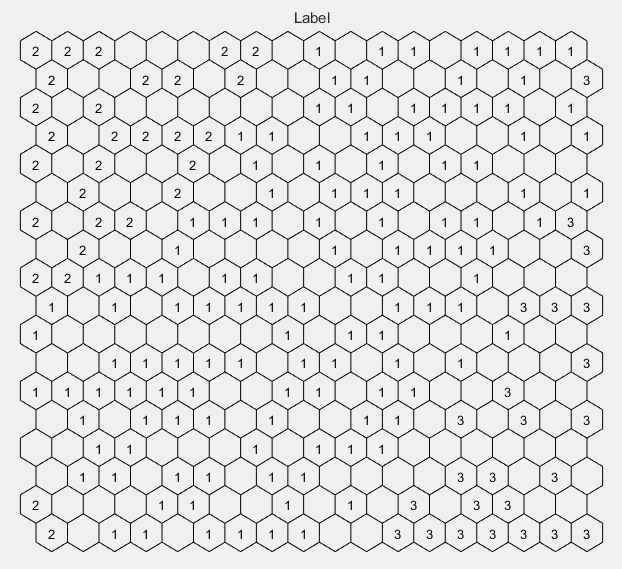
Resultados:

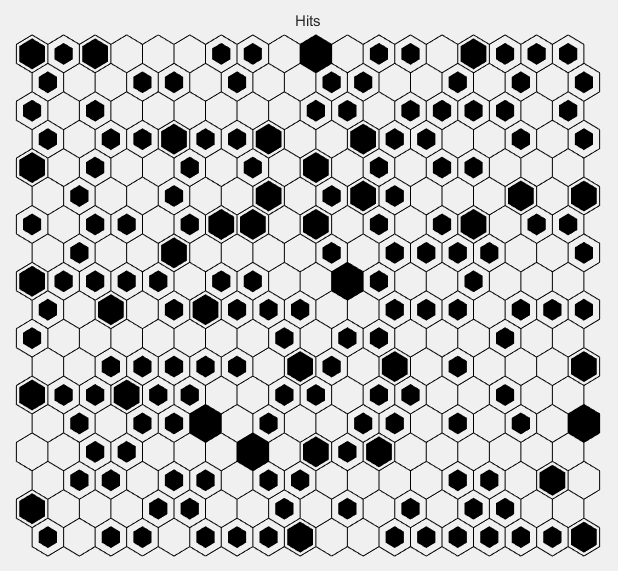
Final quantization error: 0.047

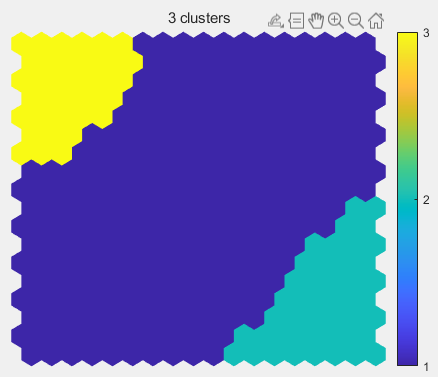
Final topographic error: 0.005

Gráficos:









Neste teste um mapa maior, de 18x18 foi testado, para verificar as influências que um mapa maior teria na visualização dos grupos. Novamente, embora as métricas de erros tenham nos dado valores menores, visualmente os mapas e gráficos se tornaram menos conclusivos.

A quantidade de processadores que não foram ativados foi muito grande, inclusive dentro dos próprios grupos. Isso faz com que os agrupamentos sejam menos reconhecíveis, o que é indesejado para uma solução por mapa de Kohonem.

**5º teste:**

Parâmetros:

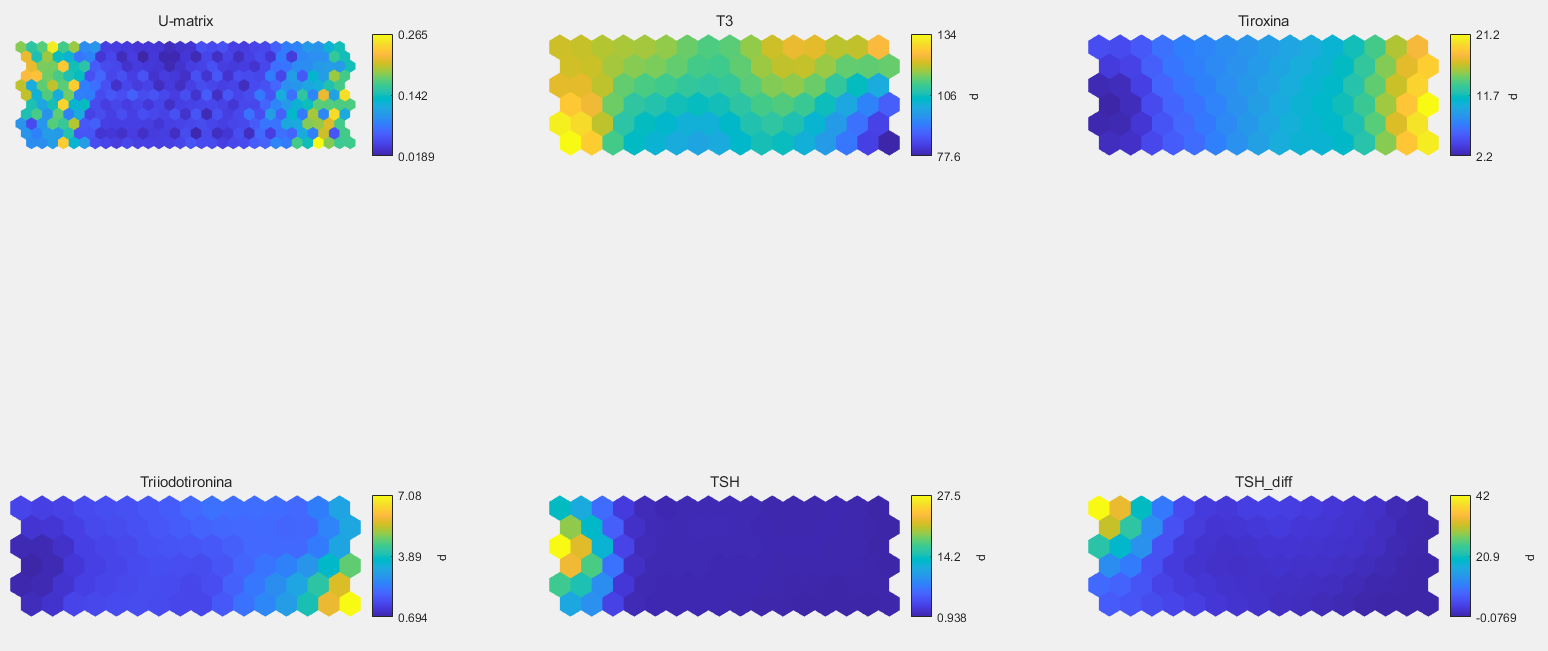
* Topologia do mapa: 6x16
* Raio: [10 1]
* Épocas na Ordenação: 1100
* Épocas no Fine Tuning: 48000
* Normalização: ‘range’

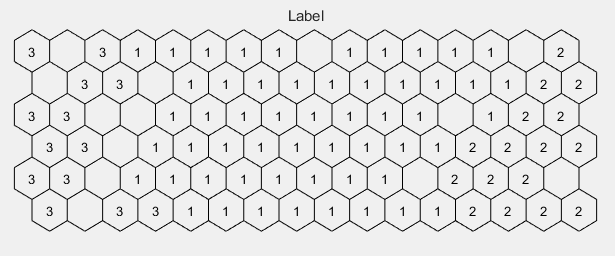
Resultados:

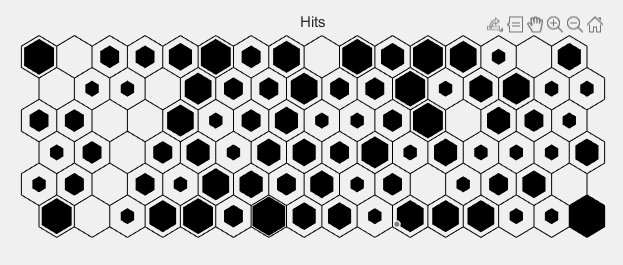
Final quantization error: 0.079

Final topographic error: 0.019

Gráficos:









Na tentativa de se conseguir grupos mais coerentes e concisos, um mapa menor foi utilizado neste teste. Observando que em todos os mapas gerados até este teste, os grupos encontravam-se em extremidades, esta topologia retangular de 6x16 foi realizada.

Novamente, o grupo 3 obteve uma boa separação, com uma divisão bem definida, porém o grupo 2 não obteve uma boa distinção.

**6º teste:**

Parâmetros:

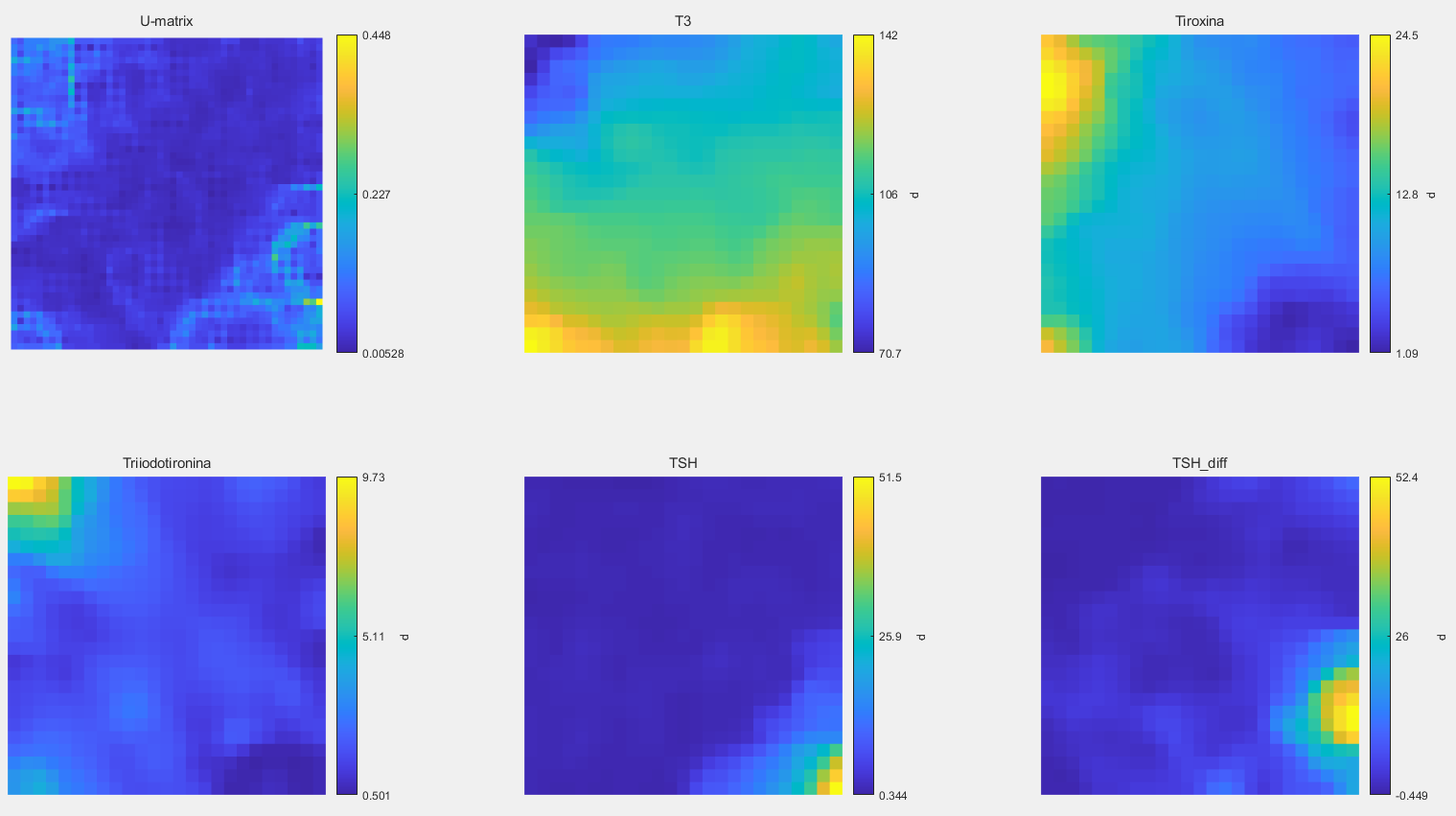
* Topologia do mapa: 25x25
* Raio: [20 1]
* Épocas na Ordenação: 1000
* Épocas no Fine Tuning: 312500
* Normalização: ‘range’

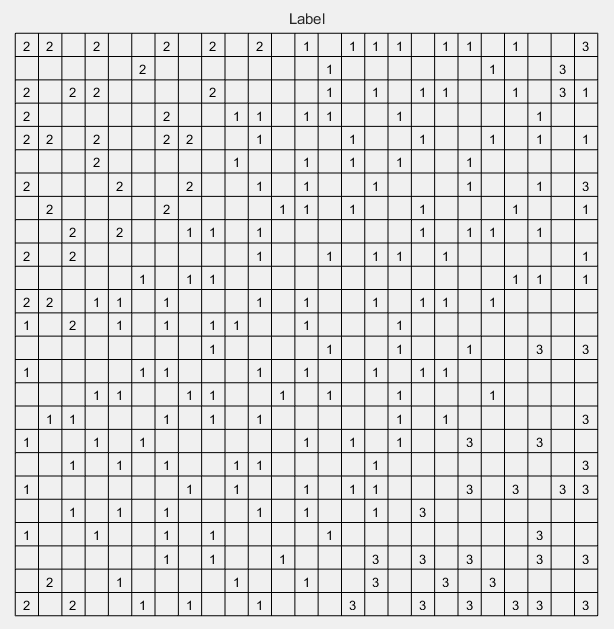
Resultados:

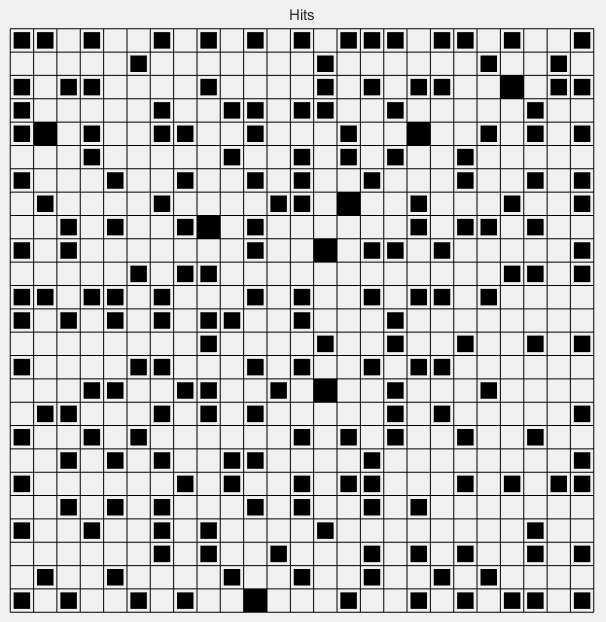
Final quantization error: 0.065

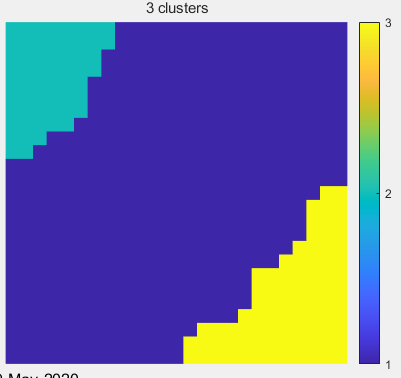
Final topographic error: 0.029

Gráficos:









Novamente um mapa “grande” foi utilizado na tentativa de alcançar uma boa separação entre os grupos. Porém, a topologia maior (25x25) gerou resultados não satisfatórios tanto em relação às métricas de erro quanto em relação à visualização dos grupos mais uma vez.

**7º teste:**

Parâmetros:

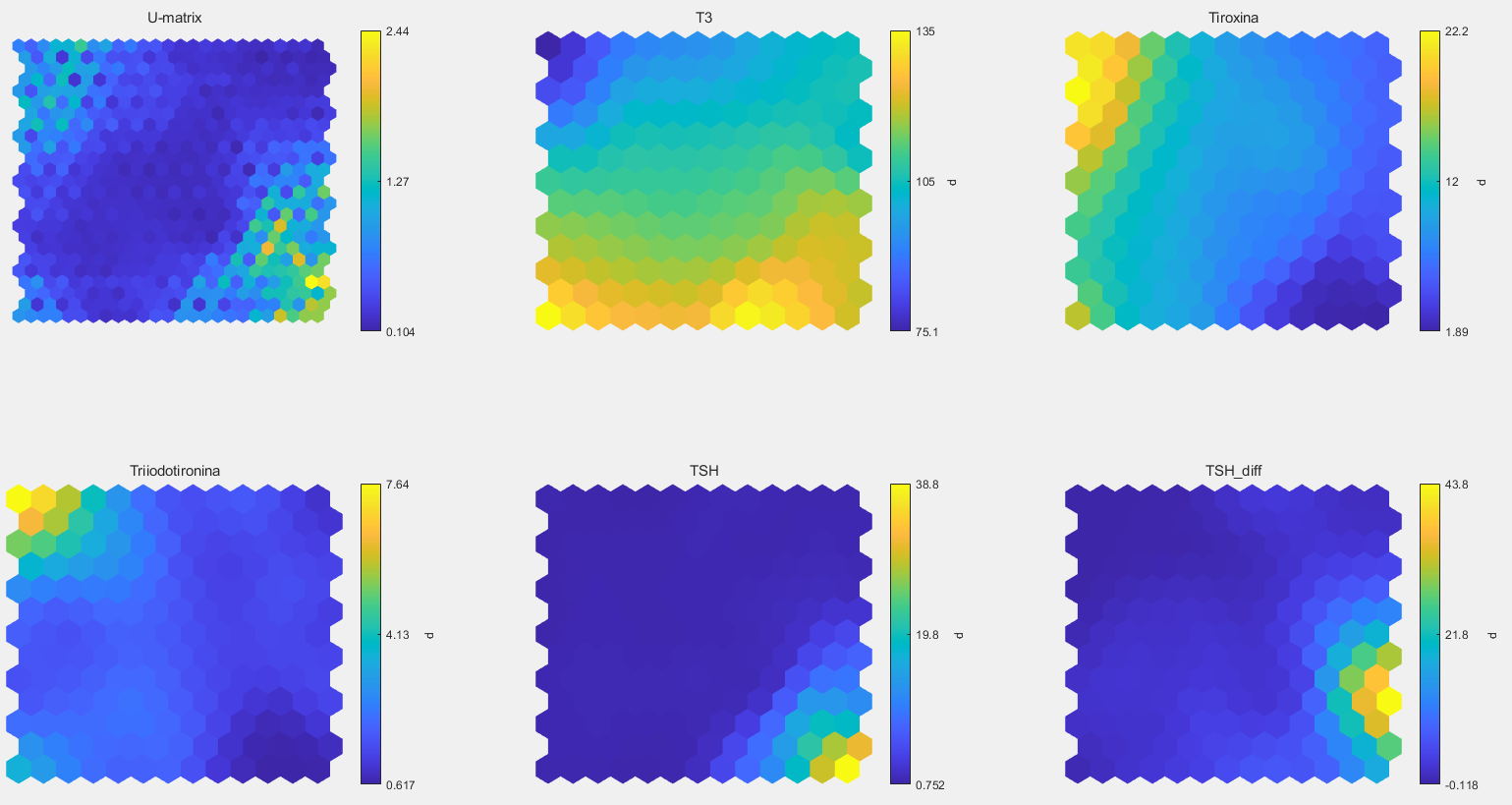
* Topologia do mapa: 13x13
* Raio: [10 1]
* Épocas na Ordenação: 1100
* Épocas no Fine Tuning: 84500
* Normalização: ‘var’

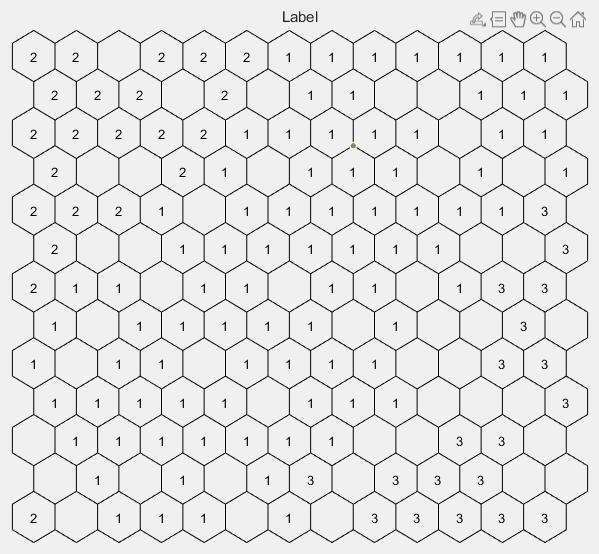
Resultados:

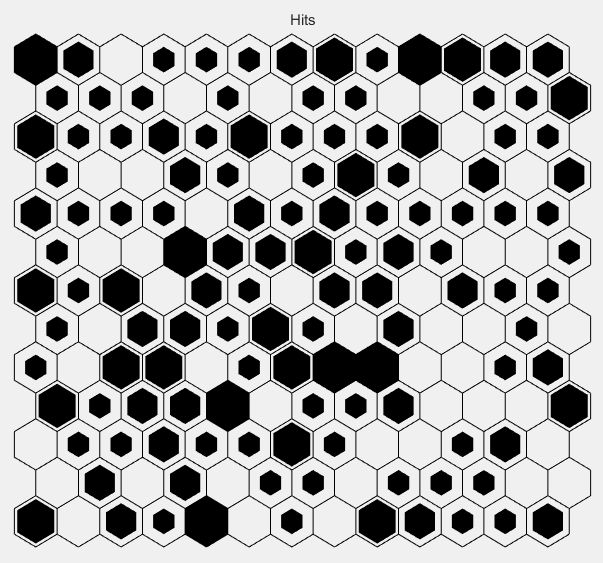
Final quantization error: 0.410

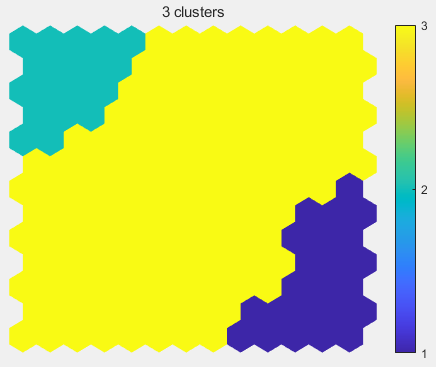
Final topographic error: 0.019

Gráficos:









Na tentativa de melhorar os resultados gerados pelo mapa 13x13 (considerado o melhor até o momento), alguns testes foram feitos modificando o tipo de normalização utilizado. No teste em questão, a normalização utilizada foi a ‘var’.

Porém, como visto nos mapas gerados, a alteração da normalização e aumento das épocas na fase de ordenação não fizeram com que o mapa ficasse mais visualmente separável. Até mesmo analisando as métricas de erro, vemos que o de quantização teve um aumento considerável.

**8º teste:**

Parâmetros:

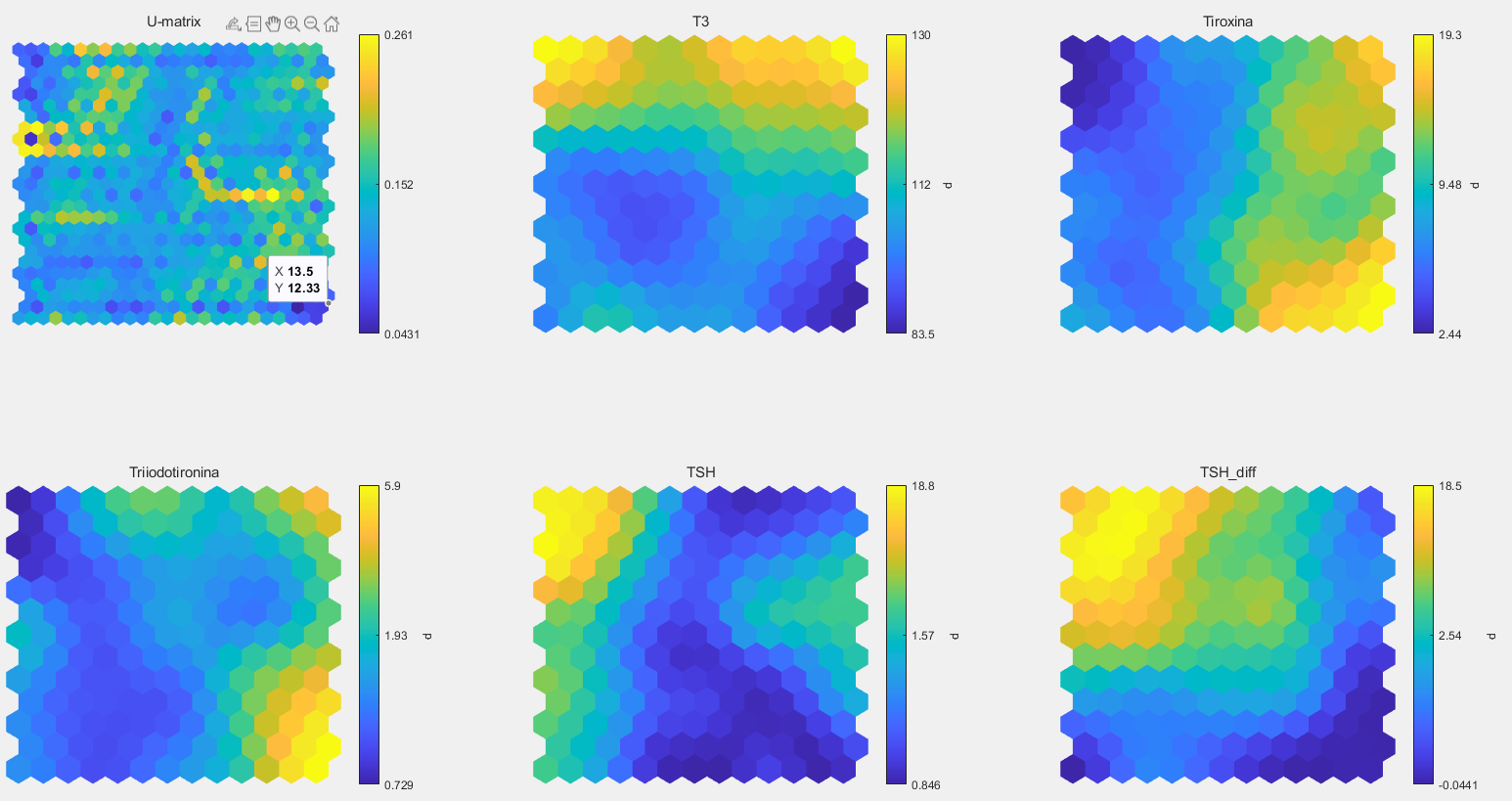
* Topologia do mapa: 13x13
* Raio: [10 1]
* Épocas na Ordenação: 1100
* Épocas no Fine Tuning: 84500
* Normalização: ‘histC’

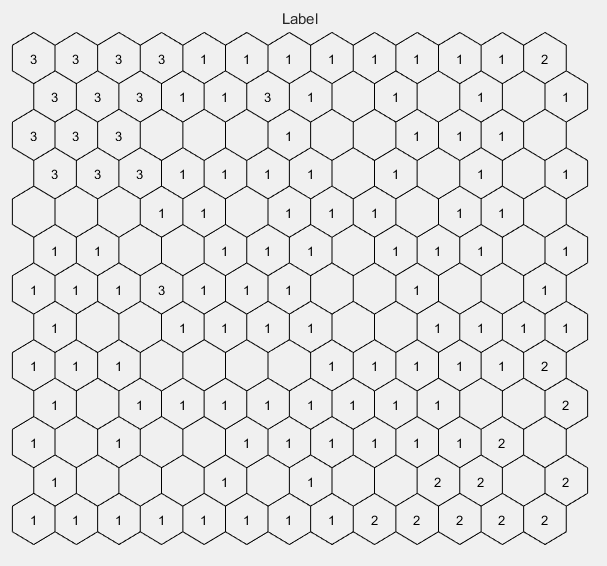
Resultados:

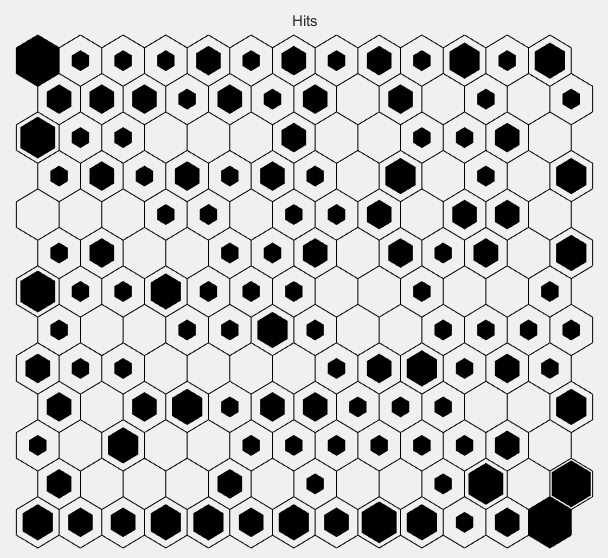
Final quantization error: 0.159

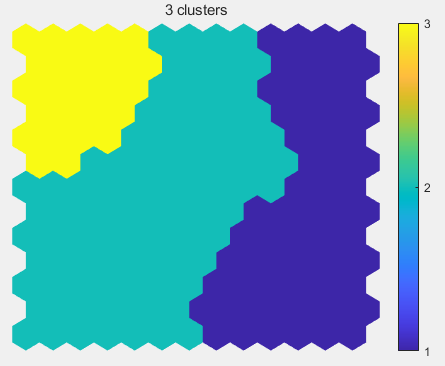
Final topographic error: 0.005

Gráficos:









Utilizando a normalização ‘histC’, observamos uma maior uniformidade nas ativações da matriz U, em geral. Porém, suas extremidades possuem “cantos” mais bem definidos como grupos de interesse. Ainda, o mapa de “hits” mostra que não existe uma borda bem delimitada entre os grupos, como aconteceu no 2º teste.

Por esse motivo, esta configuração não foi considerada a melhor. Embora em alguns aspectos tenha sido mais consistente, como nos resultados gerados pela matriz U, em outros teve um desempenho pior, como nas métricas e mapa de “hits”.

**9º teste:**

Parâmetros:

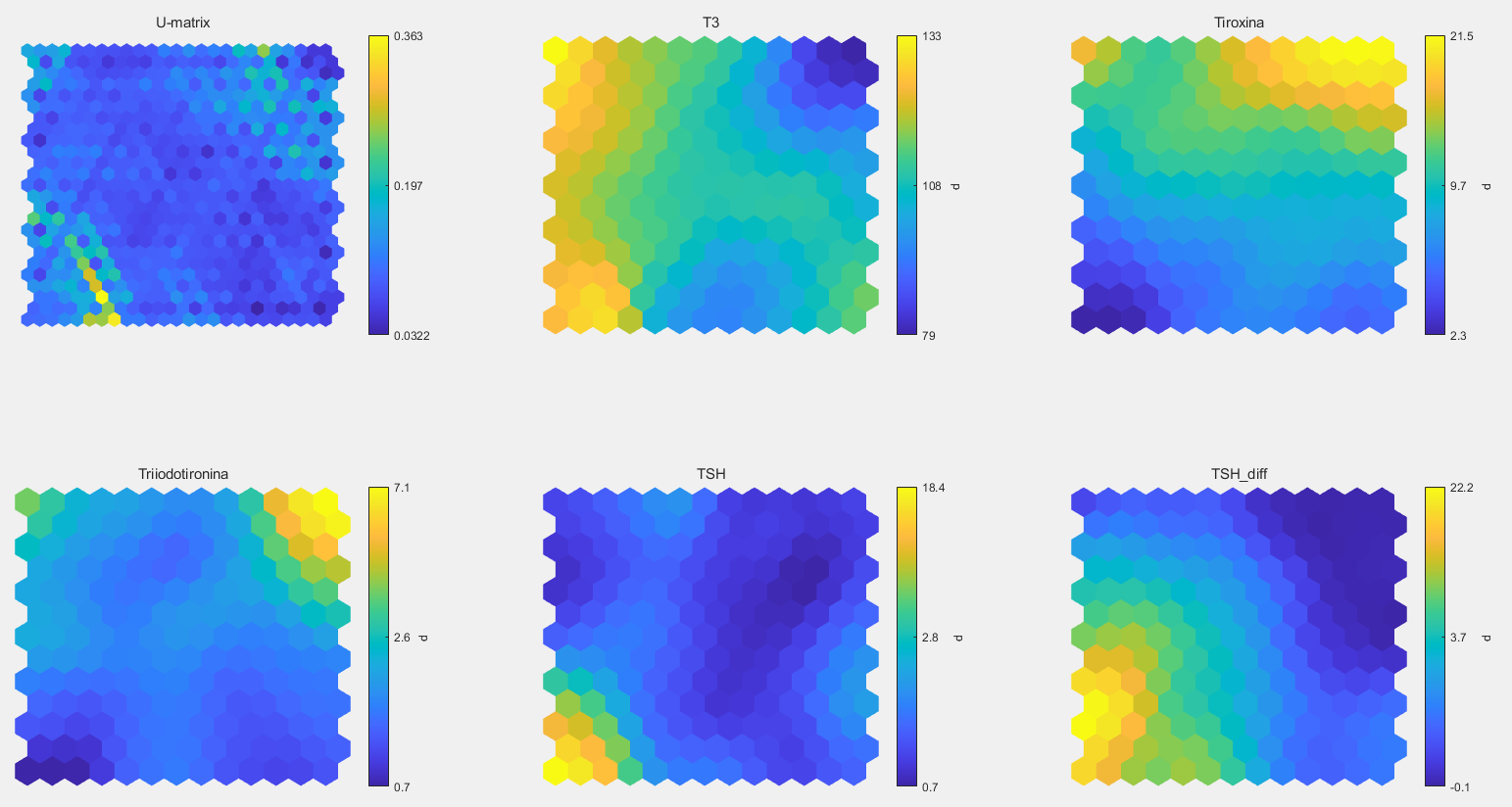
* Topologia do mapa: 13x13
* Raio: [10 1]
* Épocas na Ordenação: 1100
* Épocas no Fine Tuning: 84500
* Normalização: ‘histD’

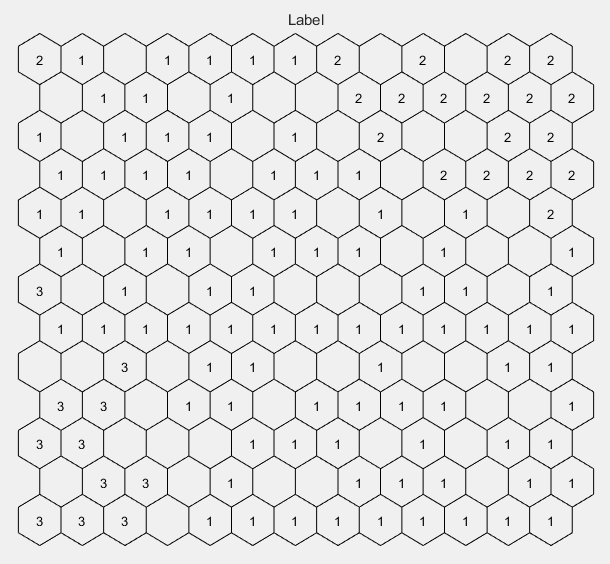
Resultados:

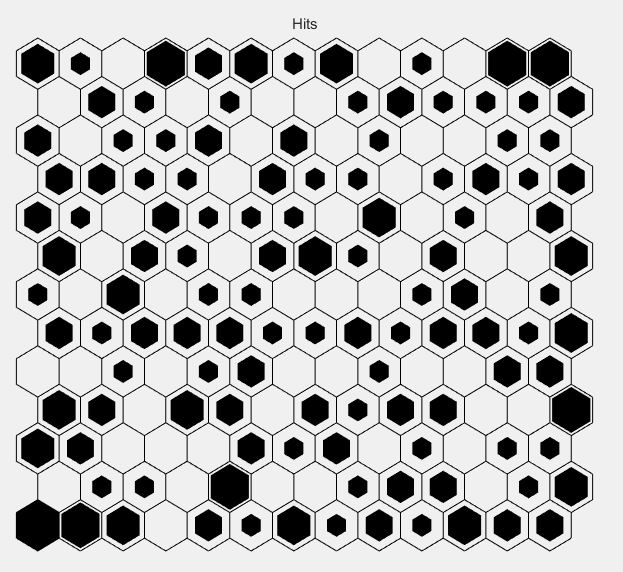
Final quantization error: 0.124

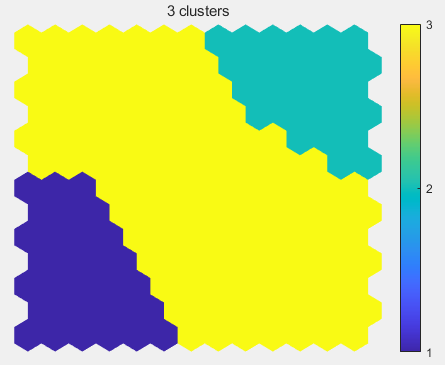
Final topographic error: 0.000

Gráficos:









Por fim, fazendo o experimento com a normalização ‘histD’, pode-se observar como a disposição das ativações na matriz U ficaram bem divididas. Nos “cantos” onde existem grupos distintos, nota-se um aglomerado mais próximo, onde, em suas bordas, existe uma área cuja distância é maior, mostrando claramente a formação de dois grupos, mais um terceiro central.

Quanto ao mapa de “hits”, percebe-se uma maior dificuldade de se separar os neurônios com ativações de todas as classes, enquanto no experimento 3, essa divisão era clara entre as classes 1 e 3.

# Conclusão

O mapa escolhido foi o da configuração de número 3, por conter uma melhor separação visual dos grupos formados principalmente no mapa de “hits”, sem minimizar totalmente a importância das métricas de avaliação de erros de quantização e topográficos. Outros gráficos, apesar de obterem métricas de erro menores, não eram visualmente tão bem interpretáveis quanto o escolhido. Talvez outros “formatos” de mapa de Kohonem poderiam ter sido testados, buscando uma melhor interpretabilidade visual dos grupos formados.

Como estamos lidando com um problema em que sabemos a quantidade esperada de grupos (eutireoidismo, hipertireoidismo e hipotireoidismo), buscamos então agrupar os resultados de forma que alcancemos essa mesma quantidade de grupos: 3. Nesse caso, podemos observar que em dois dos gráficos gerados, possuímos altos valores de Tiroxina e Triiodotironina em um mesmo “canto” dos mapas, indicando uma característica comum dos pacientes que apresentavam hipertireoidismo. Da mesma forma, em um “canto” oposto, temos altos valores de TSH e TSH\_diff, caracterizando os pacientes com hipotireoidismo. Isso nos dá uma boa direção na divisão de grupos, uma vez que os resultados fazem sentido com o problema analisado.

Além disso, na matriz U, percebemos um grupo central de neurônios mais próximos um do outro, e nos respectivos “cantos” citados acima, regiões em que os neurônios se encontram mais separados. Presume-se então, que a região central corresponde a um dos grupos existentes no problema (nesse caso o grupo de pacientes com eutireoidismo), e da mesma forma, as regiões em cantos opostos correspondem a outros grupos (hiper e hipotireoidismo) do problema.

Percebe-se também que o grupo 3 é o mais distinguível entre todos, possuindo uma região mais bem definida onde não houve ativação dos neurônios, entre ele e o grupo 1. Já o grupo 2 possui uma semelhança um pouco maior com o grupo 1, uma vez que essas regiões de separação não são tão bem definidas.

Através dos mapas visuais gerados, seria possível com a ajuda de um profissional (caso o rótulo dos dados não fosse fornecido), determinar qual grupo pertence a qual classe (hiper, hipo, eutireoidismo), discriminando as características encontradas em cada um.