Growing Hierarchical Self Organising Maps for Community Detection

David McDonald

University of Birmingham

23rd November 2016

Recap

- Using SOM to detect community structure in networks
- Most biological networks contain multi-scale (hierarchical) community structure

Growing Hierarchical SOM (GHSOM)

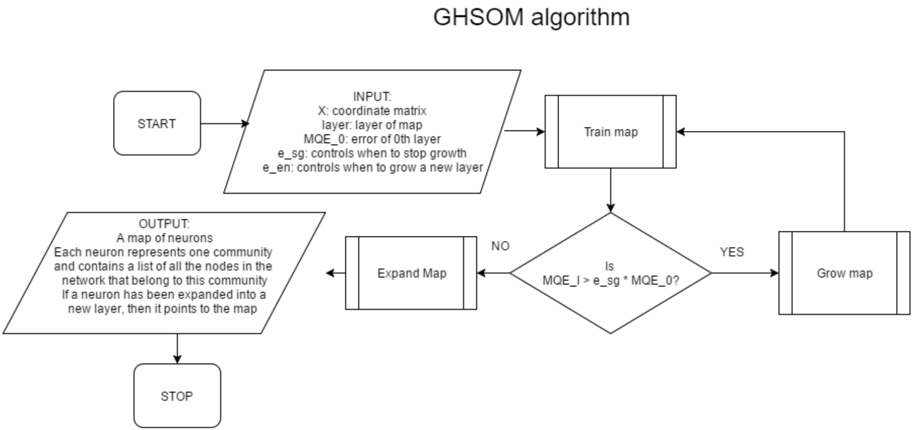
- using Growing Hierarchical SOM (GHSOM) to detect hierarchical community structure in complex networks
- a variation of SOM, proposed in [1] that can produce maps of arbitrary size and structure
- additionally, when the error of a neuron is large enough, that neuron is expanded into its own map, producing a hierarchical model
- like SOM, topological structure is preserved

Parameters

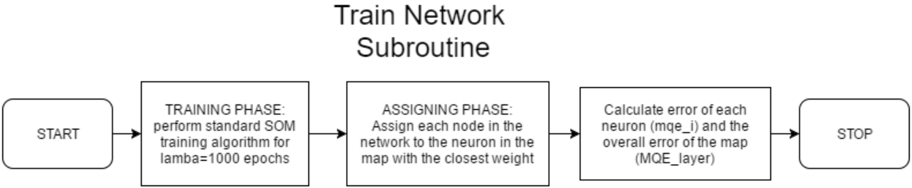
There are a few extra parameters in GHSOM than in standard SOM

- Mean Quantization Error of a map in layer l: $\mathbf{MQE}_l$
- Mean Quantization Error of a single neuron i
- Stop map growth parameter: e_{sg}
    - The error must fall below this times the previous levels error for growth to stop
    - The smaller this is the larger the map will be
- New layer parameter: e_{en}
    - If the error of a single unit is greater than this times the error of the previous layer, then a new network will be constructed using the nodes connected to this unit
    - The smaller this is, the deeper the network hierarchy

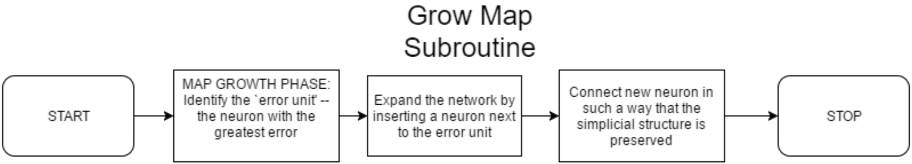
GHSOM Algorithm



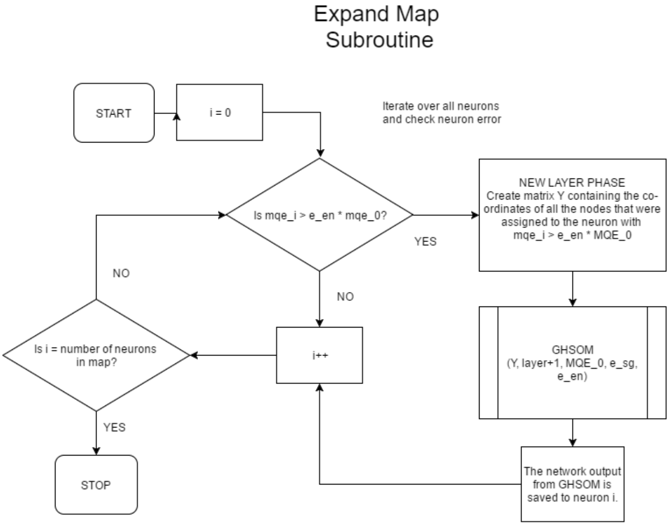
Training


Train Network Subroutine
START
TRAINING PHASE: perform standard SOM training algorithm for lamba=1000 epochs
ASSIGNING PHASE: Assign each node in the network to the neuron in the map with the closest weight
Calculate error of each neuron (mqe_i) and the overall error of the map (MQE_layer)
STOP

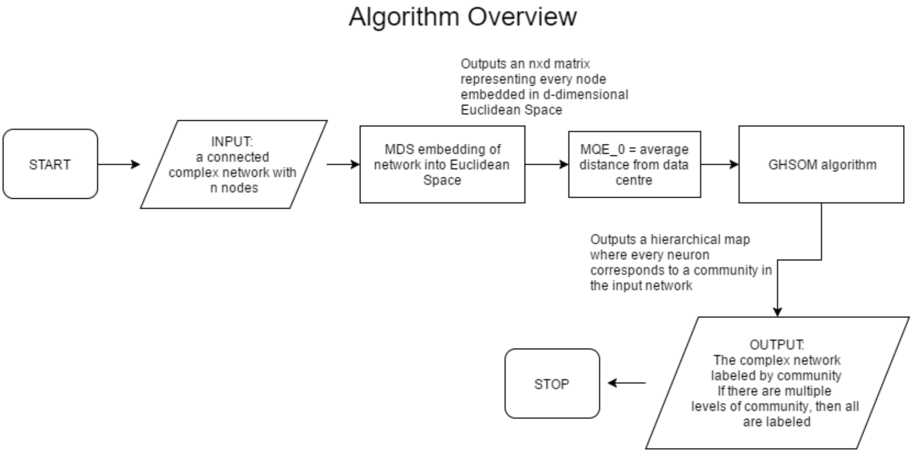
Growing the map



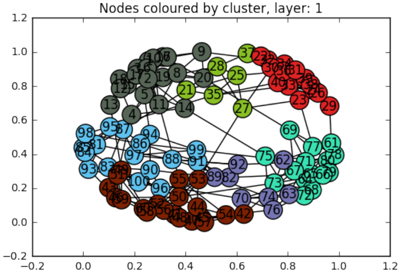
Adding a new layer

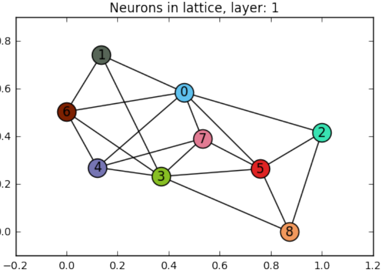


Overall Algorithm



Example on Synthetic Benchmark graph





Normalised mutual information score: 0.89584843255398483

Notice the topological information about the communities is preserved

Also, there are superfluous neurons (7)

Next Steps

- Obtain results on benchmark real world and synthetic datasets using Spearmint, Bayesian Optimisation software provided by Jasper Snoek et al. [2, 3, 4]
- Finish literature review
- Write paper (hopefully!) for 19th February submission to IJCAI 2017

What about afterwards?

- Generative Topographical Map (probabilistic counterpart of SOM)
- Latent variable learning
  - Boltzmann machine
  - Autoencoders
- Use for Active module identification

References I

Andreas Rauber, Dieter Merkl, and Michael Dittenbach.
The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data.
IEEE Transactions on Neural Networks, 13(6):1331–1341, 2002.

Michael A Gelbart, Jasper Snoek, and Ryan P Adams.
Bayesian optimization with unknown constraints.
arXiv preprint arXiv:1403.5607, 2014.

Jasper Snoek, Kevin Swersky, Richard S Zemel, and Ryan P Adams.
Input warping for bayesian optimization of non-stationary functions.
In *ICML*, pages 1674–1682, 2014.

References II

Jasper Snoek.
Bayesian Optimization and Semiparametric Models with Applications to Assistive Technology.
PhD thesis, Citeseer, 2013.