Effects of Irregular Topology in Spherical Self-Organizing Maps

Charles R. Schmidt
MS Candidate
Department of Geography
San Diego State University

May 11, 2007

Committee Members
Dr. Sergio Rey, Geography
Dr. André Skupin, Geography
Dr. Robert Malouf, Linguistics

A Thesis Proposal Presented to the Faculty of San Diego State University.

Contents

1 Introduction

The Self-Organizing Map (SOM) is an unsupervised competitive learning process developed by Teuvo Kohonen as a technique to analyze and visualize high dimensional data sets. The applications of SOM are far reaching, Kohonen (2000) provides a thorough review of the SOM literature including applications of SOM. Applications range from speech recognition and image classification to breast cancer detection and gene expression clustering. Skupin and Agarwal (2007) outline the growing interest of SOM in the GISciences, and propose that the relationship between SOM and GIScience should be bidirectional.

As explained below, the SOM is a type of artificial neural network, which facilitates a self organizing process by allowing a set of training data to arrange themselves on a network of neurons. Traditionally the SOM uses a rectangular or hexagonal grid or lattice of neurons. The topology of these grids creates a well known problem in SOM called the boundary effect. Neurons on the boundary of the hexagonal and rectangular lattices have fewer neighbors, which reduces the their ability to interact with other neurons during the self-organizing process. Using a spherical lattice is widely suggested as a solution to the problem (Ritter, 1999; Boudjemai et al., 2003; Sangole and Knopf, 2003; Wu and Takatsuka, 2006; Nishio et al., 2006). The use of the spherical lattice, however, does not completely overcome the boundary problem and the choice of which spherical topology to use for the network can be difficult to make.

With the exception of the five platonic solids, distributing points on a sphere will always result in irregular network topologies (Ritter, 1999; Harris et al., 2000). In other words, not all neurons will have the same number of neighbors. The classic method for minimizing this irregularity is to generate the spherical lattice by tessellating the sides of the icosahedron (Nishio et al., 2006). While this method will always result in a highly regular spherical topology, the main drawback is that the network size, N, grows exponentially. Other methods for arranging neurons on the sphere allow for unlimited control over network size, but yield topologies with increased irregularity (Harris et al., 2000; Wu and Takatsuka, 2005; Nishio et al., 2006). To date the literature has largely ignored the more irregular methods in favor of the tessellation based methods. A topology which yields a more flexible network size may be desirable. However, in order to address this issue of network size, we must first determine the degree to which irregularity effects the SOM.

2 Research Objectives

The objective of this research is to determine the utility of irregular spherical topologies that offer greater control over network size. Towards that end, I will develop and test new diagnostics to measure and visualize topologically induced errors in SOM. The following diagnostics will be developed and implemented:

1. Visualize the reverse quantization error by mapping neurons back on to the input data.
2. Compare internal variance of each neuron against its first-order neighborhood size.
3. Compare internal variance of each neuron against a measure of topological regularity.

These diagnostics will help facilitate the evaluation of both traditional and spherical SOMs. To satisfy the objective of this research, I will apply these diagnostics to a series of comparable SOMs. Each SOM will be trained using the same synthetic data and training parameters, but will utilize different network topologies. By formally testing for difference of means and variance in the results of the diagnostics the following questions will be addressed:

1. Does the internal variance of a neuron increase as its first-order neighborhood size, or degree, decreases?

2. Is the average internal variance of a SOM higher when a more irregular topology is used?

3 Background and Lit Review

This section is divided into four parts. The first provides a general introduction to the SOM algorithm and the problems created by using irregular topology. The second reviews the current spherical topologies used with SOMs. The third examines the flexibility of the various topologies with regard to network size. The fourth takes a look at the problem with using "uniformity" to evaluate potential topologies.

3.1 Training and the Boundary Effect

The SOM algorithm uses an artificial neural network to organize high dimensional data onto a low dimensional lattice, or map, of neurons. Each neuron contains a reference vector that models the input data. Before training, these neuron-vectors are initialized, most commonly to random values. During the training process a randomly selected observation (input vector x) searches each neuron (reference vector m_i) to find the one to which it is most similar, referred to as its Best Matching Unit (BMU c). The BMU and its neighborhood (N_c) are then adjusted to better match that observation (Kohonen, 2000). The training process is repeated a predefined number of times, or ideally until the map converges. The traditional SOM is laid out on a two dimensional plane using either a rectangular or hexagonal topology. According to Wu and Takatsuka (2006) the hexagonal structure is more uniform and generally preferred.

One drawback of building the neural lattice in a discrete Euclidean plane is the boundary of the resulting lattice. A neuron located on the boundary has fewer neighbors and thus fewer chances of being updated (Wu and Takatsuka, 2006). As observed in Figure 1, neurons in the center of the map tend to better represent the mean of the input-space. This is arguably caused by outliers being pushed to the edges of the map, where they encounter fewer competing signals.

3.2 Spherical SOM

One way to eliminate the boundary effect is to wrap the lattice around a three dimensional object such as a sphere or torus, thereby removing the edge entirely. The toroidal SOM was introduced by Li et al. (1993), however the torus is not effective for visualization, as maps generated from a torus are not very intuitive (Ito et al., 2000; Wu and Takatsuka, 2006). Ritter (1999) describes the torus as being topologically flat and suggests that a curved topology, such as that of a sphere, may better reflect directional data. A sphere also results in a more intuitive map, since we are accustomed to looking at maps based on a sphere.

Ritter (1999) first introduced the spherical SOM and several enhancements have since been suggested (Boudjemai et al., 2003; Sangole and Knopf, 2003; Nishio et al., 2006; Wu and Takatsuka, 2006). A good comparison of these enhancements can be found in Wu and Takatsuka (2006). All of these methods derive their spherical structure through the tessellation of a polyhedron as originally proposed by Ritter (1999). Wu and Takatsuka (2006) point out the importance of a uniform distribution on the sphere and that it is preferable for all neurons to have an equal number of neighbors and to be equally spaced. They find generally that the tessellation method best satisfies these conditions and specifically that the icosahedron is the best starting point (Wu and Takatsuka, 2005). Tessellation of the icosahedron results in a network of neurons, each of which have exactly six neighbors, save the original twelve which each have five. This is very close to the ideal structure in which every neuron would have exactly six neighbors. This structure has very low variances in both neuron spacing and neighborhood size.

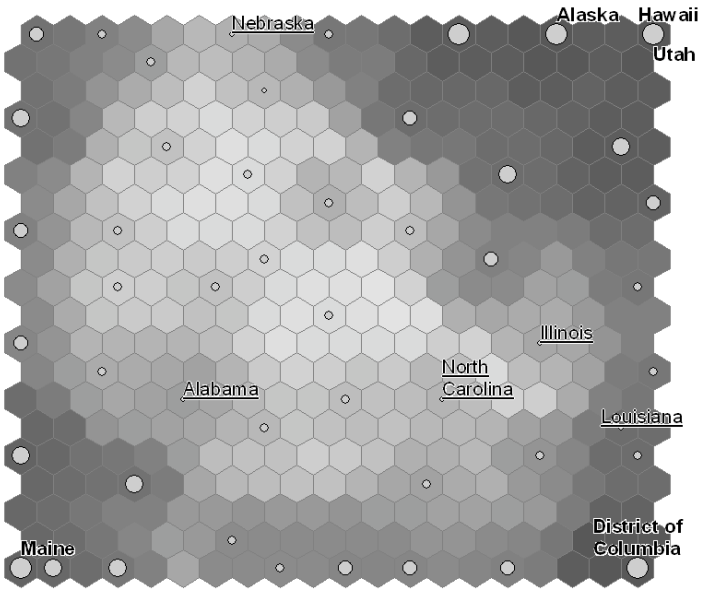

Figure 1: Fifty states plus D.C. mapped onto SOM trained with the first thirty-two census variables. Darker neurons have a relatively large differences from the mean of the states, while lighter neurons are relatively closer. Larger dots represent inputs that were poorly fit to the map, small smaller dots show inputs that were better bit to the map.

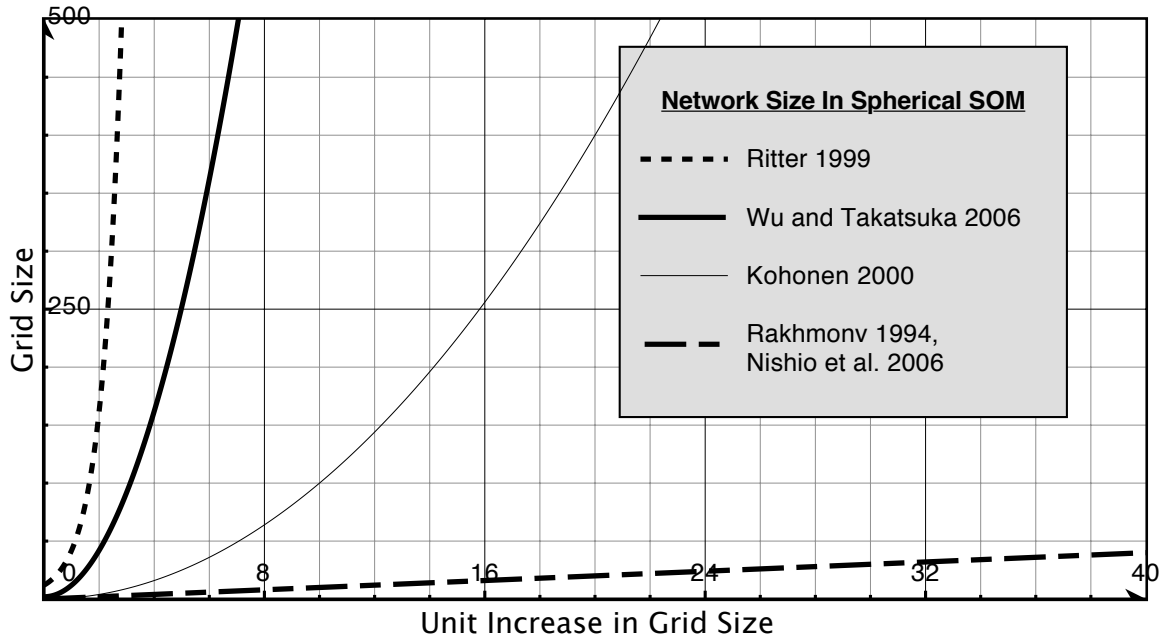

Figure 2: This figure demonstrates the achievable network size using various spherical topologies. The X axis represent a one unit increase in network size, while the Y axis represents the resulting network size. For the tessellation methods this means increasing the frequency of the tessellation by one. For the traditional Kohonen methods the dimensions are increased proportionally, such that $X = Y$

3.3 Network Size

The literature offers little theoretical guidance on network size (Cho et al., 1996). Vesanto (2005) suggests simply using a network size of $5 * \sqrt{n}$, where n is the number of observations. Given this lack of theoretical development, researchers should be cautious when using methods that limit the control of network size. As shown in Figure 2, methods for arranging an arbitrary number of points on a sphere provide a much higher degree of flexibility when choosing a network size.

The cost of relying on Ritter's tessellation method is decreased control over network size. Ritter's tessellation methods results in a network size that grows at a rate of $N = 2 + 10 * 4^f$, where f is the frequency of tessellation. Wu and Takatsuka (2006) offer a slight improvement. Rather than recursively subdividing the faces, they redivide the original icosahedron with each step resulting in $N = 2+10*f^2$. In practice 2D Euclidean SOMs also offer limited control over network size, because it is undesirable to have one dimension dramatically larger than the other. Nishio et al. (2006) try to address the issue of network size granularity by departing from the tessellation method and suggesting the use of a partitioned helix to uniformly distribute any number of neurons on a sphere. A similar method proposed by Rakhmanov et al. (1994) was dismissed by Wu and Takatsuka (2005) for failing to satisfy the uniformity conditions described above.

3.4 Uniformity

Wu and Takatsuka state that "[f]or SOM, it is desirable to have all neurons receive equal geometrical treatment" (Wu and Takatsuka, 2006, pp. 900). To satisfy this constraint two conditions must be met. Firstly, each neuron should occupy the same amount of space on the given surface. Secondly, each neuron should be bordered by the same number of surrounding neurons, and we should maximize that number. The first condition is important for visualization, but irrelevant for training. During the training of the SOM only the topology of the neurons is considered.

Based solely on measures of neuron spacing Wu and Takatsuka (2005) dismissed a method proposed by Rakhmanov et al. (1994) for distributing points on a sphere. Similarly Nishio et al. (2006) use these variance measures to support their helix algorithm for distributing points on a sphere. Table 1 shows that these metrics can be misleading and may not be comparable across topologies. The traditional rectangular and hexagonal topologies have no variance in neuron spacing and the generally preferred hexagonal structure displays greater variance in neighborhood size than the rectangular structure. The torus by comparison would have variance in neuron spacing, yet no variance in neighborhood size. The distance between two neurons is only considered during the formation of the neuronal network. At this stage the spacing is significant as it plays a part in determining neuron adjacency. However, using this measure to evaluate potential topologies for use in SOM may be misleading.

Table 1: Variances in Topologies

Topology	Grid Size	Neuron Spacing	Variance in Neighborhood Size
Rectangular	9x18	1	0.2716
Hexagonal	9x18	1	1.2138
Tessellation	162	0.25319 - 0.31287	0.0686
Rakhmanov	162	0.15779 - 0.30069	0.2908

Methods, for distributing points on the sphere, which allow for fine grained control over network size produce slightly more irregular topologies. However, no discussion of these irregularities or their effects on SOM training has occurred in the literature. Given that limited theoretical guidance is available for choosing network size the desire for finer control over the network size, should not be overlooked. In particular for a larger SOM the ideal network size may not be achievable via tessellation of the icosahedron.

4 Data and Methodology

This section is composed of two parts. The first describes the diagnostics developed for evaluating network topologies. The second describes an empirical study that will implement these methods in order to evaluate the utility of new and/or unfashionable topologies.

4.1 Diagnostics

Three diagnostics are developed to explore the effect of irregular topology on spherical SOM. The first is purely exploratory and will give the researcher a good feel for how well a SOM fits the training data. It is a simple adaptation a standard SOM diagnostic. The second diagnostic will be used the address the first of two research questions and the third diagnostic will address the second.

4.1.1 Visualize reverse mapping

The quantization error (QError) of a SOM is a simple way to measure how well data matches a trained SOM (Kohonen, 2000). The QError is measured by averaging the Euclidean distance between each observation and its best match. The QError can be decomposed to show how well each observation fits the SOM. For the first diagnostic tool I will find the Euclidean distance between each neuron and its best matching observation. The result will be a map that shows how well the neurons fit the data. This can be extended to produce a map that shows the distances of neurons to the mean (or median) of the input data.

4.1.2 Compare internal variance of each neuron against its first-order neighborhood size.

In traditional SOMs, outlying observations are pushed to the edge of the map where they encounter fewer competing signals. A prime example of this is the "Utah-Hawaii" case shown in figure 1. Relying only on the SOM one would be left to believe that the two states are similar. Upon closer inspection we see that the QError from Utah to the neuron is 1.509, the QError from Hawaii to the neuron in 1.505, but the QError from Utah to Hawaii is 3.014. In this case only Utah and Hawaii were mapped to that neuron. In a case where multiple observations land on the same neuron it is possible to measure average pairwise QErrors between those observations. This gives us a notion of internal variance for each neuron. It would be expected that in traditional SOM neurons closer to the edge will have higher internal variances. This can be extended to spherical SOM by comparing the degree of a neuron ($deg(m_i)$ or the number of adjacent neurons) to its internal variance.

Once the internal variance ($var(m_i)$) and degree ($deg(m_i)$) of each neuron have been calculated the neurons can be separated into a small number of groups based on the degree [1]. For each of these groups a variance and mean will be calculated. The expected result is that variances and means of the groups will decrease as the degree increases. This hypothesis will be tested using a ratio of variances test and a difference of means test. The result will also be visualized with a box plot.

4.1.3 Compare internal variance of each neuron against a measure of topological regularity.

The degree of each neuron can easily be calculated by taking the column sums of the first-order adjacency matrix (A). A completely regular network topology (i.e. the torus) will have no variance in these column sums. For irregular networks the variance in these column sums give us a measure of irregularity. There are many alternative ways to classify the connectivity of a network, Florax and Rey (1995) outline four such summary measures, which will be evaluated for use in this diagnostic. For each topology we can compare the internal variances as described above against a measure that summarizes the given topology's regularity.

This diagnostic is evaluated in much the same way as the last diagnostic. The internal variances are this time grouped by their topology. We can then compare the variances of internal variances and the means of the internal variances across topologies. It is expected that the distribution of internal variances will be narrower for groups trained on more regular topologies. It is further hypothesized that the means of these internal variances will decrease with more regularity. As before these assumptions will be tested using a t-test on the means and an f-test on the variances.

[1] For most topologies the number of different degrees will be limited to three or four.

4.2 Empirical Analysis

- In order to test the methods described above I will generate high dimensional synthetic data with known properties.
- I will train a SOM for each of several topologies in order to apply the diagnostics.[2]
- Knowing the properties of the training data allow us to systematically compare the diagnostics under several different topologies.
- To ensure that we can calculate an internal variance for each neuron I will generate the data so as to increase the probability that each neuron will be occupied by more than one observation.

References

Boudjemai, F., Enberg, P. B., and Postaire, J. G. (2003). Surface modeling by using self organizing maps of kohonen. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 3, pages 2418–2423 vol.3.

Cho, S., Jang, M., and Reggia, J. A. (1996). Effects of varying parameters on properties of self-organizing feature maps. *Neural Processing Letters*, V4(1):53–59.

Florax, R. J. and Rey, S. (1995). *New directions in spatial econometrics*, chapter The Impacts of Misspecified Spatial Interaction in Linear Regresion Models. Advances in Spatial Science. Springer.

Harris, J. M., Hirst, J. L., and Mossinghoff, M. J. (2000). *Combinatorics and graph theory*. Springer, New York.

Ito, M., Miyoshi, T., and Masuyama, H. (2000). The characteristics of the torus self organizing map. In *Proceedings of 6th International Conference ON Soft Computing (IIZUKA2000)*, volume A-7-2, pages pp.239–244, Iizuka, Fukuoka, Japan.

Kohonen, T. (2000). *Self-Organizing Maps*. Springer, 3 edition.

Li, X., Gasteiger, J., and Zupan, J. (1993). On the topology distortion in self-organizing feature maps. *Biological Cybernetics*, V70(2):189–198.

Nishio, H., Altaf-Ul-Amin, M., Kurokawa, K., and Kanaya, S. (2006). Spherical som and arrangment of neurons using helix on sphere. *IPSJ Digital Courier*, 2:133–137.

Rakhmanov, E. A., Saff, E. B., and Zhou, Y. M. (1994). Minimal discrete energy on the sphere. *Mathematical Research Letters*, 1:647–662.

Ritter, H. (1999). Self-organizing maps on non-euclidean spaces. In Oja, E. & Kaski, S., editor, *Kohonen Maps*, pages 97–110. Elsevier, Amsterdam.

Sangole, A. and Knopf, G. K. (2003). Visualization of randomly ordered numeric data sets using spherical self-organizing feature maps. *Computers & Graphics*, 27(6):963–976.

[2]These topologies will include the traditional hexagon and rectangle type, along with Rakhmanov et al. (1994)'s algorithm. Currently there is some uncertainty about the ability to include the Geodesic and Helix type topologies given the complexities involved with their implementation. An additional goal of this project to provide a framework on which new topologies can be easily implemented and tested by future researchers.

Skupin, A. and Agarwal, P. (2007). *(In preparation) Self-Organizing Maps: Applications in Geographic Information Science*, chapter Introduction: What is a Self-Organizing Map? In preparation for publication by Wiley.

Vesanto, J. (2005). Som toolbox: implementation of the algorithm. `http://www.cis.hut.fi/projects/somtoolbox/documentation/somalg.shtml`.

Wu, Y. and Takatsuka, M. (2005). Geodesic self-organizing map. In Erbacher, R. F., Roberts, J. C., Grohn, M. T., and Borner, K., editors, *Proc. SPIE Vol. 5669*, volume 5669 of *Visualization and Data Analysis 2005*, pages 21–30. SPIE.

Wu, Y. and Takatsuka, M. (2006). Spherical self-organizing map using efficient indexed geodesic data structure. *Neural Networks*, 19(6-7):900–910.