

High Capacity Associative Memories and Small World Networks

Neil Davey, Bruce Christianson and Rod Adams, *Member IEEE*,
Department of Computer Science
Faculty of Engineering and Information Sciences
University of Hertfordshire
College Lane, Hatfield
Hertfordshire AL10 9AB, UK
{N.Davey, B.Christianson, R.G.Adams}@herts.ac.uk

Abstract— Models of associative memory usually have full connectivity or if diluted, random symmetric connectivity. In contrast biological neural systems have predominantly local, non-symmetric connectivity. Here we investigate sparse networks of threshold units, trained with the perceptron learning rule. The units are arranged in a small world network, with short path-lengths but cliquish connectivity. The connectivity may be symmetric or non-symmetric. The results show that the small-world networks with non-symmetric weights perform well as associative memories. It is also shown that in highly dilute networks with random connectivity, it is symmetry of the weights, rather than symmetry of the connectivity matrix, that causes poor performance.

Index Terms— Associative Memory, Neural Network, Dilution, Small World Network, Weight Symmetry.

I. INTRODUCTION

It is possible to build associative memory models from networks of simple perceptrons [1, 2]. These networks perform much better than the canonical Hopfield model, both in terms of capacity and pattern completion. It is also possible to use such networks of perceptrons with sparse or diluted connectivity, and the performance is relatively robust, even at high rates of dilution [3]. Of course real neural networks have sparse connectivity, which motivates the investigation undertaken here. It is also known that in biological systems the networks have a small world characteristic [4]. That is they exhibit short path lengths between any pair of neurons, as in a random network, but also show a cliquish behaviour, with locally clustered connections. The benefit of such clustering is apparent in the mean connection length (the average wiring length), which is far smaller than in an equivalent random network. Recently Bohland and Minau [5] showed that a small world Hopfield network could be a relatively effective associative memory. This analysis is extended here using the better

performing high capacity model.

II. BACKGROUND

The simple small world model of Watts and Strogatz [4] consists of a regular N -node ring lattice. Each node is connected to $k/2$ neighbours on each side, where k is typically small compared to N . The mean path length between any pair of random points is therefore high. A fraction, p , of these local connections is then rewired to randomly selected nodes. They showed that at surprisingly low values of p , the mean path length in the network dropped dramatically, resulting in a small world regime: highly clustered but low path lengths. Many real networks have been shown to have a small world architecture, including the internet, human acquaintance networks and real networks of neurons [4]. Theoretical work [6] has now shown the detailed relationship between the characteristics of such networks.

Much is known about the effect of sparse connectivity on the standard Hopfield neural network, in particular Komlos and Paturi [7] show that the performance of the diluted Hopfield network can be characterised by the size of the first two eigenvalues of the connection matrix. Some work has also been undertaken on diluted Hopfield networks with modular and small world connectivity [5, 8]

For the higher capacity version of the Hopfield network, trained using perceptron training much less is known about the effect of connectivity patterns. It was shown in [3] that capacity falls linearly with dilution and in [9] that structured local connectivity could help in storing locally correlated data.

III. NETWORK MODEL

All the high capacity models studied here are modifications to the standard Hopfield network. The net

input, or *local field*, of a unit, is given by: $h_i = \sum_{j \neq i} w_{ij} S_j$

where $S (\pm 1)$ is the current state and w_{ij} is the weight on the connection from unit j to unit i . The dynamics of the network is given by the standard update: $S'_i = \Theta(h_i)$, where Θ is the heaviside function. Unit states may be updated synchronously or asynchronously. Here we use asynchronous, random order updates. A symmetric weight matrix and asynchronous updates ensures that the network will evolve to a fixed point. If a training pattern ξ^p is one of these fixed points then it is successfully stored, and is said to be a fundamental memory. A network state is stable if, and only if, all the local fields are of the same sign as their corresponding unit, equivalently the *aligned local fields*, $h_i S_i$, should be positive.

We examine sparse networks with both small world connectivity and random connectivity.

For the small world networks the network topology is similar to the original Watts and Strogatz model. We start with a N-ring regular lattice, with each unit connected to its k nearest neighbours, and then rewire with probability p . However a network of perceptrons is not necessarily constrained to have symmetric connections so we can generalize the network to a weighted directed graph (as is the case for real neural networks). We examine both rewiring possibilities, firstly where symmetry is maintained and secondly when it is not. In both cases the overall level of connectivity is kept equal. For purposes of comparison we also investigate networks with random, diluted connectivity, again with both symmetric and non-symmetric connections.

The networks are trained using two variants of the normal perceptron training rule. The non-symmetric learning (NSL) [1] algorithm is:

```

Begin with zero weights
Repeat until all local fields are correct
Set state of network to one of the  $\xi^p$ 
For each unit,  $i$ , in turn:
  Calculate  $h_i^p \xi_i^p$ . If this is less than T
  then change the weights to unit  $i$ 
  according to:

```

$$\forall j \neq i \quad w'_{ij} = w_{ij} + \frac{\xi_i^p \xi_j^p}{N}$$

Where ξ^p denotes the training patterns, and T is the learning threshold which here has the value of 10. All weights on removed connections are fixed at zero throughout.

And for the symmetric version (SL) the weight update is:

$$\forall j \neq i \quad w'_{ij} = w_{ij} + \frac{\xi_i^p \xi_j^p}{N} \quad w'_{ji} = w_{ji} + \frac{\xi_i^p \xi_j^p}{N}$$

Of course a network with Non-Symmetric-Connectivity (NSC) must be trained using NSL, but a symmetric network can reasonably be trained using either version of the learning rule. We denote networks with Symmetric-Connectivity (SC) graphs, trained symmetrically and non-symmetrically, as SC-SL and SC-NSL respectively. NSC-NSL denotes networks with non-symmetric connectivity and non-symmetric learning. Table 1 summarises the three types of network.

TABLE 1:
THE TYPES OF NETWORK CONNECTIVITY AND LEARNING RULE

Name	Connection Matrix	Learning Rule
NSC-NSL	Non-Symmetric	Non-Symmetric
SC-NSL	Symmetric	Non-Symmetric
SC-SL	Symmetric	Symmetric

IV. GRAPH THEORETIC QUALITY OF THE SMALL WORLD NETWORK

For the diluted, standard Hopfield model the capacity and pattern completion capability can be predicted from the relative size of the first two eigenvalues of the connection matrix [7]. The larger the difference the better the network would perform. If, for example, both eigenvalues have the same value then the network separates into two disconnected sub-graphs. In the standard fully connected Hopfield net it is the loading of the network, α , that determines the performance of the network [10]. Komlos and Paturi show that for regular connectivity graphs the parameter: $\alpha + \frac{\lambda_2}{\lambda_1}$ plays an analogous role to that of α in fully connected Hopfield networks, where λ_1 and λ_2 are the first and second largest eigenvalues of the connection matrix.

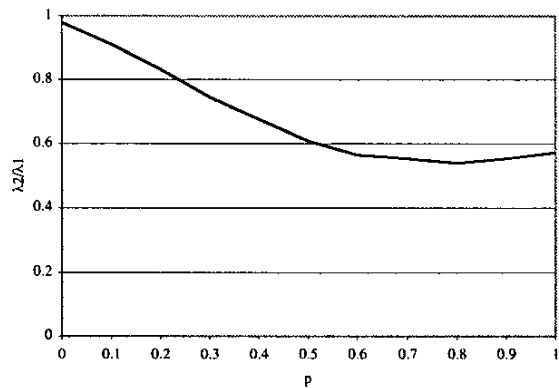


Figure 1: A 100 unit small world network, with $k = 10$. The ratio of the first two eigenvalues of the connection matrix is plotted as the rewiring probability, p , increases. Averages over 10 networks are given.

Figure 1 shows that for the locally connected network, $p = 0.0$, the first two eigenvalues are almost the same, implying that network is almost disconnected. The difference between the two eigenvalues does increase as the network is rewired, implying that the performance of the network should improve as p increases. Interestingly this improvement does not continue past $p = 0.6$.

To illustrate the potential benefits of the small world architecture, the ratio of total wiring in the small world network to the total wiring in a random network, with the same number of connections, is shown in Figure 2.

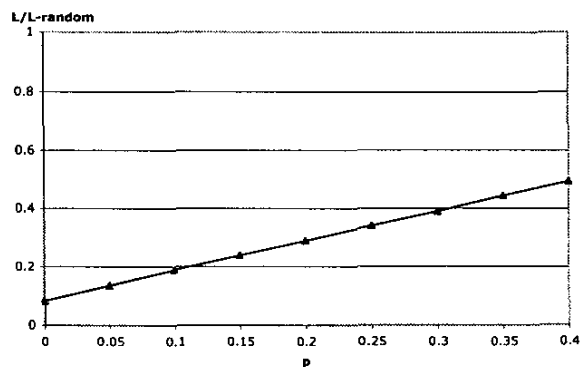


Figure 2: The fraction of total wiring when comparing the 500, unit $k = 38$ small world networks with random networks with the same number of total connections

V. PERFORMANCE MEASURE

We are interested in how well the small world networks and random networks, trained using the various perceptron style learning rules described above, perform as associative memories. The capacity of such networks is determined by the number of incoming connections (K) that each perceptron has. For random pattern sets a perceptron can learn up to $2K$ patterns [10]. Assuming roughly regular connectivity graphs (as is the case here) the capacity will be determined by the level of dilution and not the specific pattern of connections, and hence is not subject to empirical investigation.

We use, R , the normalised mean radius of the basins of attraction, as a measure of attractor performance in these networks. It is defined as:

$$R = \left\langle \left\langle \frac{1 - m_0}{1 - m_1} \right\rangle \right\rangle$$

where m_0 is the minimum overlap an initial state must have with a fundamental memory for the network to converge on that fundamental memory, and m_1 is the largest overlap of the initial state with the rest of the fundamental memories. The angled braces denote a double average over sets of training patterns and initial states. Details of the

algorithm used can be found in [3]. A value of $R = 1$ implies perfect performance and a value of $R = 0$ implies no pattern correction.

For small world networks a further refinement of this basic measure is of interest. With predominantly local connections it is particularly difficult for a small world network to correct patterns in which the errors are contiguous [11]. We therefore also measure $R_{\text{contiguous}}$ in which all initial states contain only contiguous patterns of error.

The final measure we report is the symmetry of the

trained weights. This is evaluated as: $\sigma = \frac{\sum_{i,j} w_{ij} w_{ji}}{\sum_{i,j} w_{ij}^2}$. For a

symmetric matrix this takes the value +1. For an anti-symmetric matrix it takes the value -1 and for a random set of weights it will be roughly zero.

VI. RESULTS

A. Small World Regime

Here 500 unit networks were arranged in a ring, as described earlier. Experiments were conducted using two different network configurations. Firstly initializing the network with each unit connected to its $k = 38$ nearest neighbours, and a training set of 12 random, unbiased patterns. The second configuration had $k = 76$ and a training set of 24 patterns. Each network type was then rewired with varying rewiring probabilities p . For the networks with non-symmetric connectivity, NSC, the rewiring was non-symmetric. For both network types with symmetric connectivity, symmetry was maintained in the rewiring. However in all cases the total number of connections was maintained across network types, for a specific value of p . As described earlier it is interesting to measure the performance of these networks at pattern correction for both contiguous and non-contiguous domains of error: both R and $R_{\text{contiguous}}$ are reported. Figure 3 shows the R values for the $k = 38$ networks.

The first point of interest is the poor performance of the networks with only local connectivity, $p = 0$. This should not be surprising – as seen in Section III such networks are almost disconnected. As p increases so does the performance of the networks with non-symmetric weights. When $p = 0.4$ both NSL networks are performing perfectly. The network with symmetric weights, SC-SL however performs poorly across the full range of p values.

Figure 4 shows the performance of the networks at correcting contiguous domains of error. Up to $p = 0.3$ all the networks perform very poorly. Then first NSC-NSL and later SC-NSL show a rapid jump in performance. At $p = 0.4$ NSC-NSL is correcting even these patterns perfectly

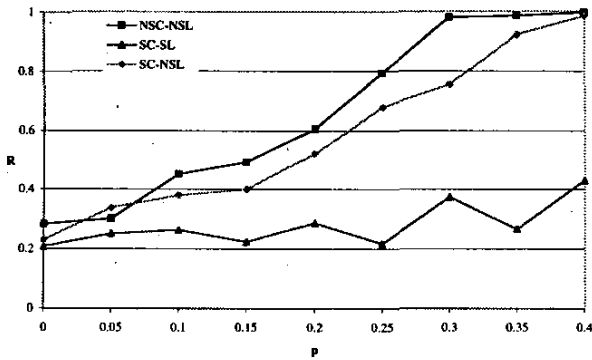


Figure 3: 500 unit small world network, $k = 38$ and a training set of 12 patterns. Results are averages over 5 runs at each value of the rewiring parameter, p . R values of the three different types of network are shown.

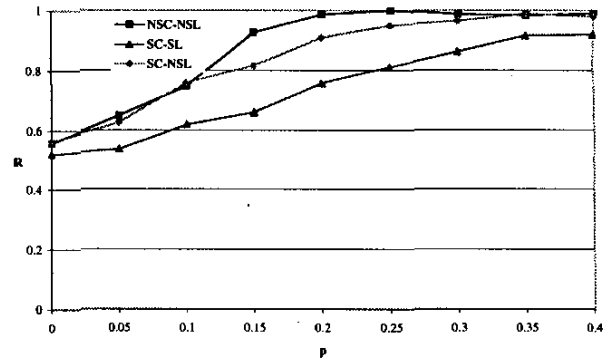


Figure 5: 500 unit small world network, $k = 76$ and a training set of 24 patterns. Results are averages over 5 runs at each value of the rewiring parameter, p . R values of the three different types of network are shown.

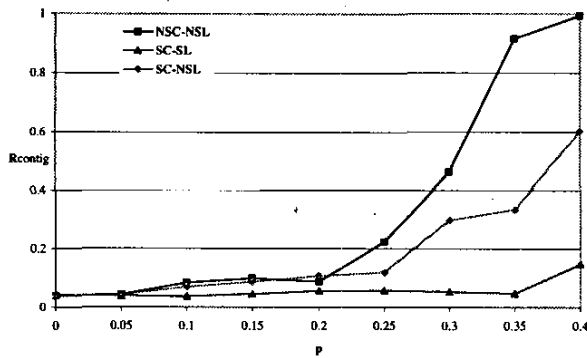


Figure 4: 500 unit small world network, $k = 38$ and a training set of 12 patterns. Results are averages over 5 runs at each value of the rewiring parameter, p . R_{contig} values of the three different types of network are shown.

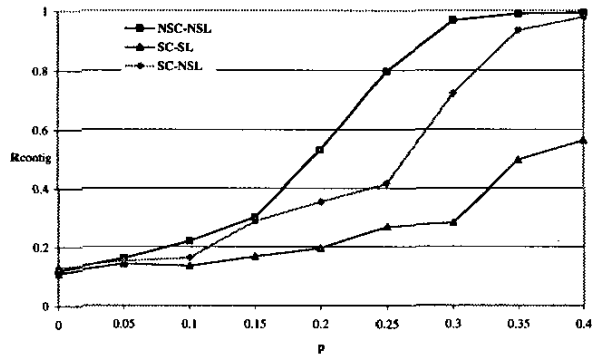


Figure 6: 500 unit small world network, $k = 76$ and a training set of 24 patterns. Results are averages over 5 runs at each value of the rewiring parameter, p . R_{contig} values of the three different types of network are shown.

Figures 5 and 6 show the results when both the loading and level of connectivity are doubled, so that 24 patterns are stored in networks with $k = 76$. The general pattern of performance is repeated, with the pattern correction ability of the non-symmetric learning networks both correcting contiguous errors perfectly with $p = 0.4$. The performance difference between SL and NSL is however not as great in this case, suggesting that higher levels of connectivity are not as detrimental to SL.

B. Random Connectivity

The results in the previous section suggest that at high rates of dilution, symmetric weights do not give good performance, but symmetric connectivity is not a problem. To investigate this phenomena more thoroughly a set of experiments were performed on the simpler randomly connected diluted networks.

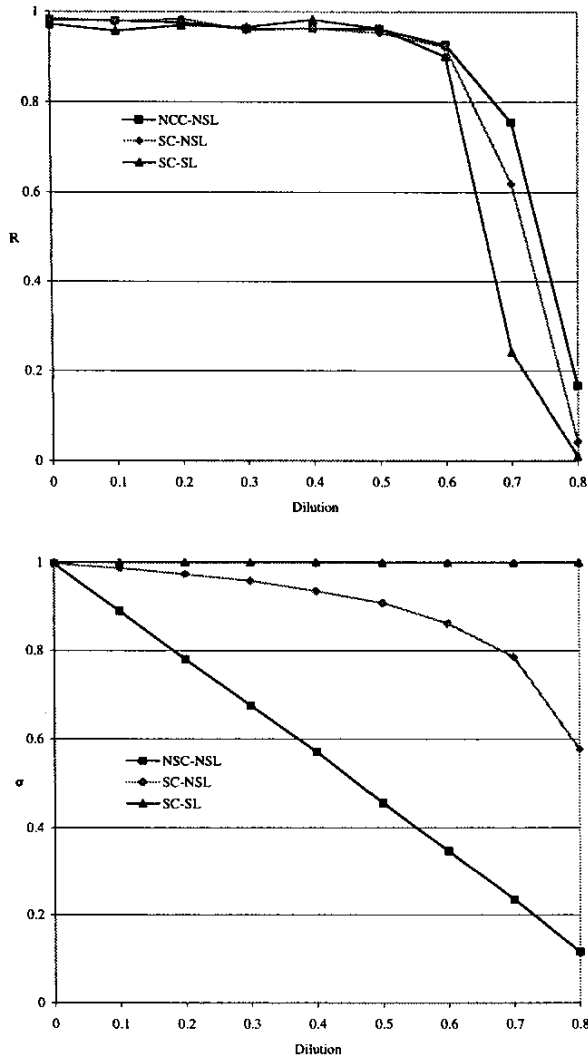


Figure 7: Random diluted networks, each network is 100 units and is trained with 10 random patterns. Results are averages of 50 runs at each rate of dilution. Above the R values and below the weight symmetry.

Figures 7 and 8 show the performance of the three types of diluted network described earlier. Note that for random networks there will be no difference between the normal R measure and $R_{\text{contiguous}}$. At both loadings R values eventually decrease as dilution increases. At high rates of dilution it is apparent that SC-SL networks have much lower R values than the other two types. It is interesting that SC-NSL and NSC-NSL have similar performance, suggesting that it is the imposition of symmetric weights in SC-SL that causes the relatively poor performance at high dilution rates.

It is known [3] that for *fully* connected networks NSL produces very symmetric weights. However the symmetry results show that both non-symmetric learning rules produce increasingly less symmetric weights as the dilution

increases. It is also noteworthy that the NSC-NSL network still performs well with low levels of weight symmetry (for example the 10 training pattern, 0.6 dilution, R value of 0.907).

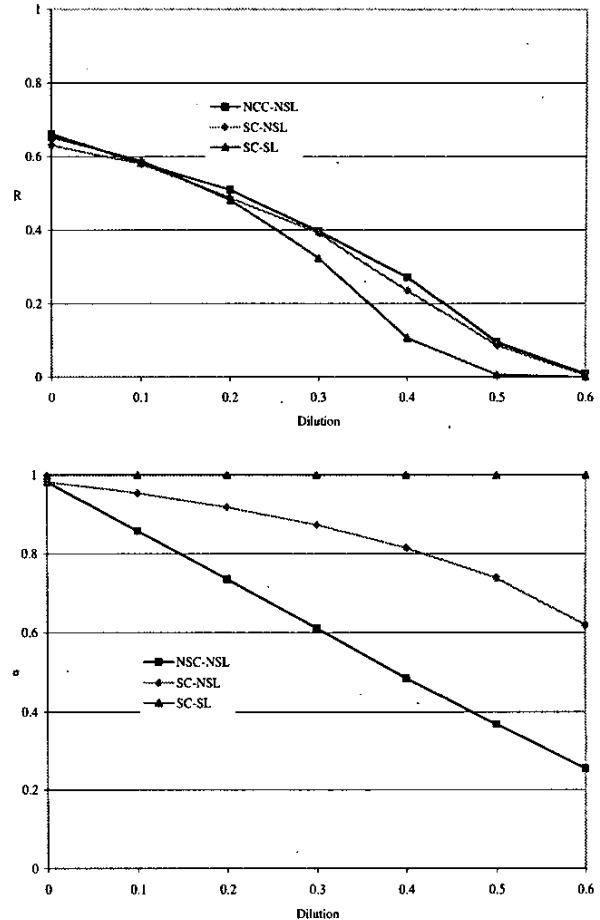


Figure 8: Random diluted networks, each network is 100 units and is trained with 30 random patterns. Results are averages of 50 runs at each rate of dilution. Above the R values and below the weight symmetry.

VII DISCUSSION

Associative memories built from small world networks are attractive because they have lower mean wiring cost than an equivalent random network. They also correspond more closely to real biological neural networks. It has been shown here that small world networks of perceptrons can perform very well with a suitable level of rewiring, and are even capable of correcting contiguous patterns of error. The exceptions to this result were small world networks that adhered to strict weight symmetry, namely symmetric connectivity and symmetric weights. However the constraint of symmetric

connectivity alone did not hamper the networks performance in a major way.

In order to investigate the role of symmetry more fully networks with non-structured dilute connectivity were examined. These networks confirmed the results from the small world networks. At very high levels of dilution the performance of the strictly-symmetric network was much worse than that of either of the other non-symmetric models. Interestingly the symmetrically connected networks with the non-symmetric learning rule became progressively less symmetric as dilution increased.

These results also raise a more general question about the role of symmetry in this type of model. Of course in the canonical Hopfield network the weights are symmetric, and this is sufficient to guarantee point attractors under asynchronous dynamics [10]. Here the networks without a symmetry constraint become highly non-symmetric at high rates of dilution but the convergence behavior is not disturbed. Further theoretical or empirical work is needed to explain the dynamic behaviour of these perceptron based networks with non-symmetric weights. Another issue that we have not been able to pursue, as yet, is the relationship between the connectivity level, the loading of the network, and the value of rewiring parameter, p , at which good attractor performance is restored.

Ultimately the most interesting question that this type of model raises is: what is the optimal pattern of connectivity to maximize performance and minimize wiring cost? The answer to this could throw light on the connectivity pattern in real neural networks.

[11] A. J. Noest, "Domains in Neural Networks with Restricted-Range Interactions," *Physical Review Letters*, vol. 63, 1989.

REFERENCES

- [1] S. Diederich and M. Opper, "Learning of Correlated Patterns in Spin-Glass Networks by Local Learning Rules," *Physical Review Letters*, vol. 58, pp. 949-952, 1987.
- [2] B. M. Forrest, "Content-addressability and learning in neural networks," *Journal of Physics A*, vol. 21, pp. 245-255, 1988.
- [3] N. Davey and R. Adams, "High Capacity Associative Memories and Connection Constraints," *Connection Science*, 2004.
- [4] D. Watts and S. Strogatz, "Collective Dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440-442, 1998.
- [5] J. Bohland and A. Minai, "Efficient Associative Memory Using Small-World Architecture," *Neurocomputing*, vol. 38-40, pp. 489-496, 2001.
- [6] M. E. J. Newman, "Models of the Small World," *Journal of Statistical Physics*, vol. 101, pp. 819-841, 2000.
- [7] J. Komlos and R. Paturi, "Effect of Connectivity in an Associative Memory Model," *Journal of Computer and System Sciences*, vol. 47, pp. 350-373, 1993.
- [8] D. O'Kane and A. Treves, "Short- and long-range connections in autoassociative memory," *Journal of Physics A: Mathematical and General*, vol. 25, pp. 5055-5069, 1992.
- [9] S. P. Turvey, "Analysing and Enhancing the Performance of Associative Memory Architectures," in *Computer Science*. Hatfield: University of Hertfordshire, 2003.
- [10] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*. Redwood City, CA: Addison-Wesley Publishing Company, 1991.