Connectivity Graphs and the Performance of Sparse Associative Memory Models

Weiliang Chen, Rod Adams, Lee Calcraft, Volker Steuber and Neil Davey

Abstract—**This paper investigates the relationship between network connectivity and associative memory performance using high capacity associative memory models with different types of sparse networks. We found that the clustering of the network, measured by Clustering Coefficient and Local Efficiency, have a strong linear correlation to its performance as an associative memory. This result is important since a purely static measure of network connectivity appears to determine an important dynamic property of the network.**

I. INTRODUCTION

THERE are many problems to be overcome before artificial neural networks can be built that resemble the mammalian cortex. Not the least of these problems is finding a way to connect the neurons so that the network functions well, and is physically realizable. In particular natural cortical systems are very sparsely connected — in the mouse cortex only approximately 1 in 100 million of all possible connections are actually made. Yet at the level of an individual neuron connectivity is very high with roughly 10000 incoming and outgoing connections being made. Of course the connectivity in such systems is not random. In fact the connectivity of the system will attempt to meet two competing objectives. Firstly the amount of fiber used overall will be minimized; connecting fiber is in several senses expensive: it creates heat that must be dissipated, it needs constant resource replenishment and it needs physical space. Secondly information must be spread widely in the cortex for its integration and for global computation to take place; and this would appear to entail much distal connectivity.

In recent years scientists have successfully introduced measures from graph theory into the investigation [1-5]. Most of them report the cortex to be a so-called "small world" network, which has a short path length similar to a random network, and a high clustering property similar to a locally connected network [2]. Some further research investigates the effects of network connectivity as associative memory performance, suggesting that the connectivity of a network indeed affects the performance significantly [6, 7]. However, a conclusive relationship still has not been revealed.

In previous experiments we discovered that one of the connectivity measures, the Local Efficiency of a network, has a strong correlation with its associative memory performance [6]. This finding inspired us to investigate how the connectivity of a sparse associative memory affects its

performance. Different connectivity measures were used in different network types, including networks with more biological plausibility such as a network with Gaussian distribution of connections, and networks with modular structure.

II. MEASURES OF THE NETWORK CONNECTIVITY

A. Path Length and Clustering Coefficient

Watts and Strogatz [2] investigated a series of real world networks and discovered that these networks were neither completely regular nor completely random. Graph theoretical measures were used to qualify the properties associated with their connectivity. In particular, two measures, the mean Path Length (L), and the Clustering Coefficient (C), were introduced.

The *Path Length* is the minimum number of arc traversals to get for one node to another. An average over all pairs of vertices is used to produce $L(G)$ for a graph G. Denoting the length of the shortest path for each pair of vertices as d_{ij}, the *Path Length* of a graph G with N vertices is

$$L(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} d_{ij}$$

It is notable that for a disconnected graph, $L(G)$ is problematic since d_{ij} for any pair of disconnected vertices is undefined.

The *Clustering Coefficient* $C(G)$ of a graph G is defined as follows. Firstly, define C_i, the local clustering coefficient of node i, as

$$C_i = \frac{\text{\# of edges in } G_i}{\text{maximum possible \# of edges in } G_i} = \frac{\text{\# of edges in } G_i}{k_i(k_i - 1)}$$

where G_i is the subgraph of neighbours of i (excluding i itself), and k_i is the number of neighbours of vertex i. C_i denotes the fraction of every possible edge of G_i which actually exist. The *Clustering Coefficient* of G, $C(G)$, is then defined as the average of C_i over all vertices i of G:

$$C(G) = \frac{1}{N} \sum_{i \in G} C_i$$

Figure 1 gives a simple example of the calculation of Path Length and Clustering Coefficient.

Science and Technology Research Institute, University of Hertfordshire, College Lane, Hatfield, Herts, AL10 9AB, United Kingdom
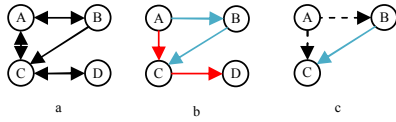w.3.chen, r.g.adams, l.calcraft, v.steuber, n.davey@herts.ac.uk

Figure 1. Example of Path Length, Clustering Coefficient of a directed graph. A,B,C,D: nodes; Solid line: existing connections. Arrows: directions of the connections. Graph a: The whole graph. Graph b: Subgraph of path from A to D. The red path (A-C-D) takes 2 steps and is therefore the shortest one (A-B-C-D takes 3 steps), so d_{AD} = 2. The mean Path Length of this graph is (d_{AB} + d_{AC} + d_{AD} + d_{BC} + d_{BD} + d_{CD} + d_{BA} + d_{CA} + d_{DA} + d_{CB} + d_{DB} + d_{DC}) / 12 = 18 / 12 = 3/2. Graph c: For the Clustering Coefficient, we measure the subgraph of A's neighbours (B & C), There are two possible edges (B-to-C and C-to-B) but only one exists, so C_A is 1/2. Dashed line: connections from A. Consequently, $C(G) = (C_A + C_B + C_C + C_D) / 4 = 3/8$.

It is found [2] that a locally connected network has both high mean *Path Length* and high *Clustering Coefficient*. On the other hand, a random network has both low mean *Path Length* and low *Clustering Coefficient*. Between these two extreme cases there are a large number of networks which have a low mean *Path Length* like the locally connected network (the so-called small-world effect), as well as a high *Clustering Coefficient*. This characteristic turns out to be a common feature in real networks. Examples of such networks are real neural networks (the cat's cerebral cortex, the neural network of *C.elegans*), social networks and the *World Wide Web* [2, 5, 8].

B. Global and Local Efficiency

Watts and Strogatz [2] characterize the Path Length and the Clustering Coefficient as two different measures. These measures in fact can be unified, as shown by Latora and Marchiori [4], to one single measure, the *efficiency* of a network, as well as its subnetworks.

For a directed graph G (connected or disconnected), the average *efficiency* $E(G)$ is defined by the following formula:

$$E(\mathrm{G}) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}$$

In particular, the *efficiency* of a fully connected network, which contains all *N(N-1)* edges, is named as $E(\mathrm{G}^{ideal})$. For a topological, directed graph, $E(\mathrm{G}^{ideal}) = 1$. Unlike the mean *Path Length*, $E(\mathrm{G})$ will not be divergent for a disconnected graph because $1/d_{ij}$ is defined as 0 for any disconnected pair *i,j*.

To formalize the *Path Length* and the *Clustering Coefficient* to a single measure, two new terms, the *global efficiency* and the *local efficiency* are introduced. The *global efficiency* of a graph G, E_{glob}, is defined as

$$E_{glob} = \frac{E(\mathrm{G})}{E(\mathrm{G}^{ideal})}$$

In fact E can be calculated for any subgraph of G. Therefore the local properties of G can be characterized by the *local efficiency*, E_{loc}, which is the average efficiency of each node's neighbour subgraphs,

$$E_{loc} = 1/N \sum_{i \in G} \frac{E(\mathrm{G}_i)}{E(\mathrm{G}_i^{ideal})}$$

G_i is defined as the subgraph of all the neighbours of vertex *i*. As before G_i^{ideal} is the ideal case of G_i which contains all possible edges. The *small-world network* is now characterized as a set of networks with both high *global* and *local efficiency*. Figure 2 is an example of calculating the Global and Local Efficiency.
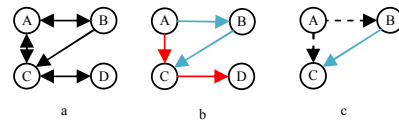

Figure 2. Example of Global and Local Efficiency of a directed graph. A,B,C,D: nodes; Solid line: existing connections. Graph a: The whole graph. Graph b: Subgraph of path from A to D. The red path (A-C-D) takes 2 steps which is the shortest one (A-B-C-D takes 3 steps), so d_{AD} = 2. The Global Efficiency of this graph is ($1/d_{AB}$ + $1/d_{AC}$ + $1/d_{AD}$ + $1/d_{BC}$ + $1/d_{BD}$ + $1/d_{CD}$ + $1/d_{BA}$ + $1/d_{CA}$ + $1/d_{DA}$ + $1/d_{CB}$ + $1/d_{DB}$ + $1/d_{DC}$) / 12 = 7/9. Graph c: For the Local Efficiency, we measure the subgraph of A's neighbours (B & C), so $E(\mathrm{G}_C) = (1/d_{BC} + 1/d_{CB}) / 2 = 1/2$. Dashed line: connections from A. Consequently, $E_{loc} = (E(\mathrm{G}_A) + E(\mathrm{G}_B) + E(\mathrm{G}_C) + E(\mathrm{G}_D)) / 4 = 3/8$. Note: For a large sparse network the Clustering Coefficient and the Local Efficiency are usually not the same, see [4] for details.

III. THE CONNECTIVITY OF THE REAL MAMMALIAN CORTEX

Braitenberg and Schüz [9] investigated the connectivity of the mammalian cerebral cortex and suggested a system with two levels of connectivity. At a high level, the network is constructed mainly from area-to-area excitatory connections between pyramidal cells. At low level, the network within an area is constructed from short range excitatory and inhibitory connections of both pyramidal and non-pyramidal cells.

Much research [5, 8, 10] indicates that the area-to-area connectivity has a low *Path Length* but high *Clustering Coefficient* (high *global* and *local efficiency*), just as in a *small-world network*. On the level of individual neurons, the connectivity is so complex that only some general statistics and hypotheses can be produced [9]. One important hypothesis [11] suggests that the basic functional unit of the mammalian cortex is the "minicolumn", a columnar structure constructed from several hundreds of neurons. Although this hypothesis is still debatable [12], it suggests that the network

of an associative memory model could be constructed as a set of inter-connected modules.

IV. THE HIGH CAPACITY ASSOCIATIVE MEMORY MODEL

A. Dynamics

The units in the network are simple bipolar threshold devices, summing their inputs and firing according to the threshold. The net input, or *local field,* of a unit, is defined 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 update rule of network dynamics is slightly different from the one used in the canonical model

$$S'_i = \begin{cases} 1 & \text{if } h_i > \theta \\ -1 & \text{if } h_i < -\theta \\ S_i & \text{for other cases} \end{cases}$$

where S'_i is the new state of S_i, and θ is the update threshold of the dynamics.

Unit states may be updated synchronously or asynchronously. The asynchronous update as well as a symmetric weight matrix guarantees the network will evolve to a fixed point. However, we found that without these restrictions, the network could still achieve fairly similar convergence properties. In our experiment we used asynchronous update with $\theta = 0$ for simplification.

If a trained pattern $\boldsymbol{\xi}^u$ is one of the fixed points of the network then it is successfully stored and is called a *fundamental memory*.

B. Learning

A one-shot Hebbian training is commonly used as the standard learning rule of the Hopfield Net. Although simple to implement and also statistically tractable, this learning rule has several drawbacks. The one-shot Hebbian rule does not guarantee that all trained patterns are actually learnt (which means they may not be *fundamental memories*). Furthermore it is widely known that such a network has quite a low theoretical maximum capacity ($0.14N$ for a fully connected network with *N* units[13]). In the standard Hopfield model the connections are required to be symmetrical, which is certainly not the case in real biological networks. Therefore our networks have not symmetric constraint. The performance of an associative memory can be improved using other classes of learning rules [13]. In our experiments, we adopted and modified Gardner's perceptron learning rule [14] which guarantees all trained patterns will be memorized, as well as giving a significantly higher theoretical maximum capacity of up to $2N$ for unbiased patterns. The detailed training process is given as follows:

```
Denoting T as the learning threshold
Begin with a zero weight matrix
```

```
Repeat until all units are correct
 Set the state of the network to one of the ξᵖ
 For each unit, i, in turn:
   Calculate its local field hᵢᵖ

   If (ξᵢᵖhᵢᵖ < T ) then change the weight on
connections
   into unit i according to:
```

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

```
   where {Cᵢⱼ} is the connection matrix
 End For
End
```

C. Performance Measure

It is important to investigate not only the capacity of the associative memory model but also the ability of *fundamental memories* to act as *attractors* in the state space of the network dynamics.

To measure this we use the *Effective Capacity* of the network, *EC* [3, 15]. The *Effective Capacity* of a network is a measure of the maximum number of patterns that can be stored in the network with *reasonable* pattern correction still taking place. We take a fairly arbitrary definition of *reasonable* as correcting the addition of 60% uniform noise to within an overlap of 95% with the original fundamental memory. Varying these figures gives differing values for *EC* but the values with these settings are robust for comparison purposes (see [15] for the effect on *Effective Capacity* of varying the degree of applied noise, and the required degree of pattern completion). For large fully-connected networks the *EC* value is about 0.1 of the maximum theoretical capacity of the network, but for networks with sparse, structured connectivity *EC* is dependent upon the actual connection matrix *C*.

The *Effective Capacity* of a network is defined as follows:

```
Initialise the number of patterns, P, to 0
Repeat
  Increment P
  Create a training set of P random patterns
  Train the network
  For each pattern in the training set
    Degrade the pattern randomly by adding 60% of
    noise With this noisy pattern as start state,
    allow the network to converge
      Calculate the overlap of the final
        network state with the original pattern
  End For
  Calculate the mean pattern overlap over all
final states
Until the mean pattern overlap is less than 95%
The Effective Capacity is P-1
```

V. Models Examined

Four different types of sparse networks were examined. These were two non-modular networks and two modular ones. The first non-modular network is the well-studied Watts-Strogatz small-world network [2]. The second non-modular network is a network with Gaussian-distributed connectivity. The third is a modular network which is initialized from fully connected modules and then rewired externally at different rewire rates. The final model is a modular network constructed by creating modules with Gaussian-distributed connectivity internally and random intermodular connections.

The connectivity properties of a network were measured using Global Efficiency, Clustering Coefficient, and Local Efficiency. The first series of networks had 5000 units (N = 5000) and 249 connections per unit (k = 249); this value of k is chosen so that in a modular network of 250 units it is fully connected. The second series had 5000 units (N = 5000) and 499 connections per unit (k = 499). Experiments were repeated 20 times and a mean value was reported.

A. Watts-Strogatz Small-world Network

This model followed Watts and Strogatz's original idea [10]. N units were arranged on a one dimensional ring. Each unit was initially connected from its k nearest units. A fraction q denoted the proportion of these connections which were randomly rewired. Particularly, q = 0 gave a locally connected network and q = 1 constructed a random network. In the experiments, q was increased from 0 to 0.5 by a step of 0.05, and then from 0.5 to 1 by a step of 0.1. This was due to the fact that the performance of the network increased significantly at low q and tends to saturate when q exceeded 0.5. Figure 3 gives the transformation of a network from regular local (q = 0) to Small World (q = 0.1) then to random (q = 1).
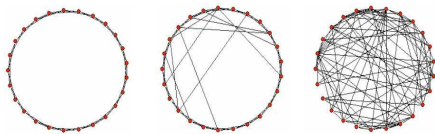

Figure 3. The W-S model [2]. Left: A lattice or locally connected network (q = 0). Middle: A small-world network with rewiring q = 0.1. Right: A random network (q = 1). In all three cases the number of afferent connections is, k = 4. Diagrams generated with the Pajek package [16]. The left network has both high L and C, whilst the right network has both low L and C. The middle one has low L but high C (L: mean Path Length; C: Clustering Coefficient).

B. Gaussian Distributed Network

In the mammalian cortex most of the connections are local, with the probability of any two neurons in the same area being connected, falling off in a Gaussian like manner [17] (also see Figure 4). This was the main inspiration for our *Gaussian Distributed Network*. In this model, all units were still arranged on a one dimensional ring as in the *W-S* network. However, the connections were constructed according to a Gaussian distribution of distance between connected units. The standard deviation, σ, was varied to get different distributions of connections. By increasing σ, the network changed from a strongly locally connected network to an almost randomly-connected network.
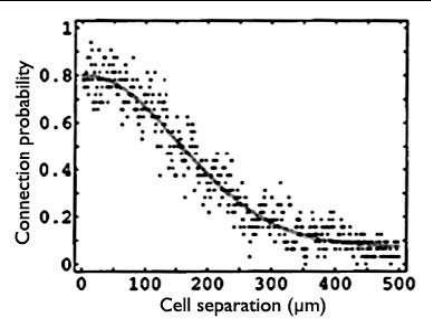

Figure 4. The probability of a connection between any pair of neurons in layer 3 of the rat visual cortex against cell separation. Taken from [17], with permission .

C. Fully-Connected Modular network

In this model the postulated columnar structure of the mammalian cortex [11] was adopted. The network initially contained m internally fully connected networks, defined as *modules*. At the beginning there were no interconnection between the modules. Thus it can be treated as m fully connected associative memories. The network was then connected by rewiring the intramodular connections to random connections anywhere in the whole network. A fraction p denoted the proportion of rewired connections (Figure 5).
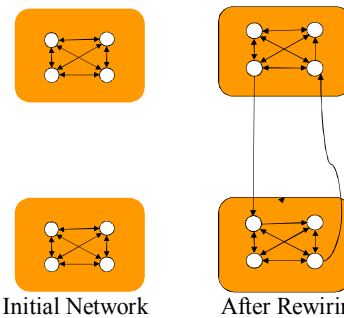

Initial Network   After Rewiring

Figure 5. The construction of a fully modular network. The network was initialized as several discrete modules (left), and then gained intermodular connections by rewiring the intramodular connections. Note that the regularity of the network is maintained during the rewiring (each node always has 3 incoming connections).

In the experiments the number of modules, m, was defined as $N / (k+1)$, so that the modules could be fully connected to keep the same degree of connectivity as the other models. Therefore for the network with k = 249, m = 20. And for the network with k = 499, m = 10. For simplification we denote this network as the *Modular Network* in later sections.

D. Gaussian Distributed Modular Network

The final model examined was the most complex one of the four models in this paper. This model was defined by two levels of connectivity. The connections of a unit were

classified as intramoduar connections (define *intra-k* as the number of intramodular connections per unit) and intermodular connections (define *inter-k* as the number of intermodular connections per unit). At the intramodular level, the connections were constructed using a Gaussian distribution, characterised by the standard deviation σ. At the intermodular level, the connections were connected randomly. Although the proportions of intra/inter modular connections varied, the total number of connections per unit was maintained, that is, *intra-k + inter-k = k* for all networks. We denote this network as the *Gaussian Modular Network* followed by *intra-k* and *inter-k*. For example, a network with *intra-k* = 49 and *inter-k* = 200 is denoted as "Gaussian Modular 49 200".

VI. RESULTS

A. General Results from Each Model

Here we give individual results for each type of network based on their wiring strategies. For simplification only results for 5000N, 249k networks are presented. The results for 5000N, 499k networks will be summarized at the end of this section.

The first result is from the Watt-Strogatz network. Figure 6 gives the relationship between Effective Capacity and rewiring rate of the network. The Effective Capacity increases rapidly from $q = 0$ to $q \approx 0.6$ and then saturates.
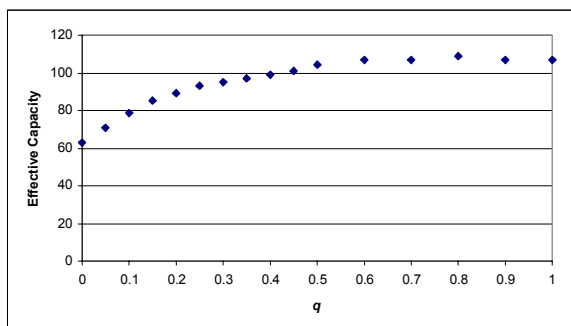

Figure 6. Effective Capacity against rewiring rate *q* in W-S network. *N* = 5000, *k*=249. The Effective Capacity increases untill *q* ≈ 0.6 and saturates later.
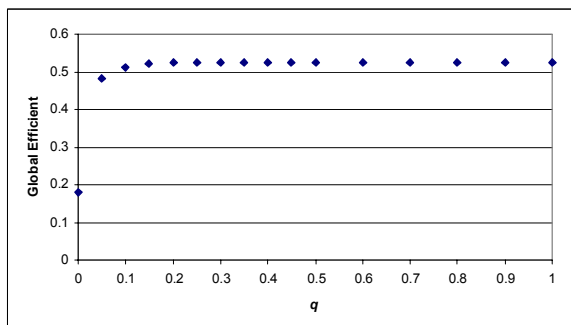

Figure 7. Global Efficiency against rewiring rate *q* in W-S network. *N* = 5000, *k*=249. The Global Efficiency increases rapidly untill *q* ≈ 0.2 and saturates in all others.

Figure 7 and 8 show the way that Global Efficiency and Clustering Coefficient vary with the rewiring of the network. Global Efficiency saturates very quickly, much more quickly than the Effective Capacity. However the Clustering Coefficient declines less rapidly and appears to have an inverse relationship with Effective Capacity. The Local Efficiency shows a similar pattern as the Clustering Coefficient (Figure 9). These correspondences are investigated further in the next section.
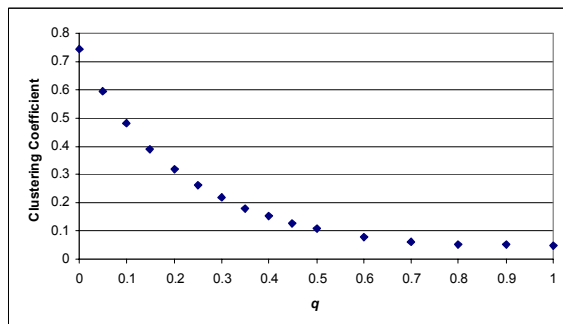

Figure 8. Clustering Coefficient against rewiring rate *q* in W-S network. *N* = 5000, *k*=249. The Clustering Coefficient decreases untill *q* ≈ 0.6 and saturates later. Interestingly it has the same saturating point as Figure 6 (*q* ≈ 0.6).

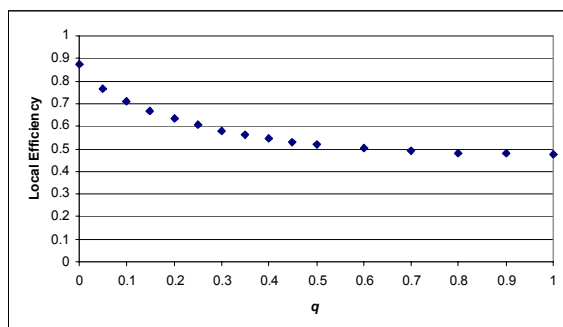Figure 9. Local Efficiency against rewiring rate *q* in W-S network. *N* = 5000, *k*=249. The Local Efficiency decreases rapidly untill *q* ≈ 0.6 and saturates later. Again it has the same saturating point as Figure 6 (*q* ≈ 0.6).

For the Gaussian network we plot the Effective Capacity against the standard deviation of the connection distribution, σ (Figure 10). The Effective Capacity increases with σ and reach a saturation value of about 110 when σ is 1000. Note that the maximum Effective Capacity of the Gaussian network and the W-S network are the same. This is because at the right hand extreme of both figures we have randomly connected networks.
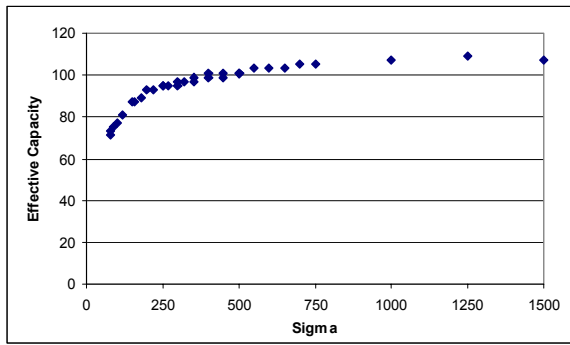
Figure 10. Effective Capacity against *sigma* in Gaussian network. *N*: 5000, *k*=249. The Effective Capacity increases untill *sigma* is about 1000 and saturates later.

The results for the *Modular Network* are shown in Figure 11. Since the *Modular Network* starts from discrete modules and is then rewired into a random network, its Effective Capacity is initially lower than that of the previous two models. But with rewiring it approaches the maximum value obtained in the other two models.
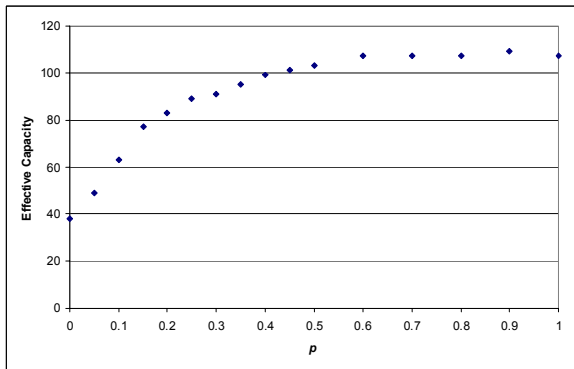

Figure 11. Effective Capacity against rewiring rate *p* in the Modular network. *N* = 5000, *k*=249. The result is similar to the one in the W-S network (Figure 6), despite the difference of initial value.

The final network we investigated is the Gaussian Modular network. The results are shown in Figure 12. Interestingly the Effective Capacity does not change very much as the intraconnections are made less local. The Effective Capacity is quite high throughout whatever the value of σ.
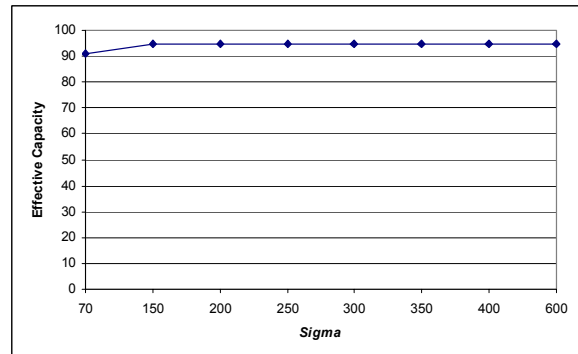

Figure 12. Effective Capacity against *sigma* of intraconnections distribution in Gaussian Modular network. *N* = 5000, *intra-k* = 199,*inter-k* = 50. There is slight change when change the distribution within modules but not significant.

B. Effective Capacity and Clustering

The above results suggest an interesting hypothesis: the performance of associative memory models, measured by the Effective Capacity, is clearly determined by some measures of the network connectivity, such as clustering and the efficiency of local sub-network. Therefore here we plot the Effective Capacity against each connectivity measure in all four types of networks. Figure 13 shows the relationship between Effective Capacity and Global Efficiency in six different networks. Obviously there is no simple relationship between the two measures. However, Figure 14 gives our more significant result. It shows that, in all six networks there is a linear relationship between Effective Capacity and Clustering Coefficient. Moreover this relationship is independent of the detailed topology of the network. Figure 15 gives the linear regression for this data. The *R-Square* measure is 0.99, so the fit is highly linear.
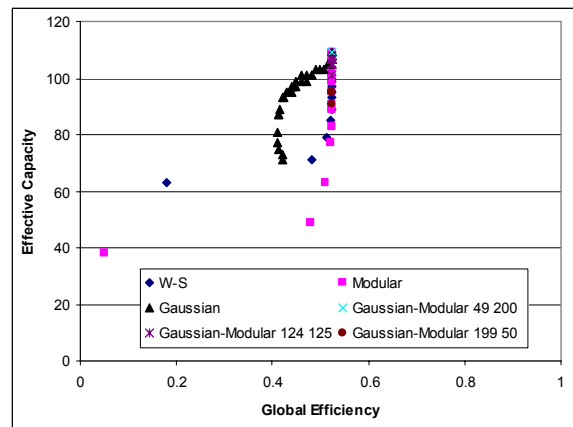

Figure 13. Effective Capacity against Global Efficiency. *N* = 5000, *k*=249. Results from four different types of networks are plotted together. No clear relationship can be found in this figure.
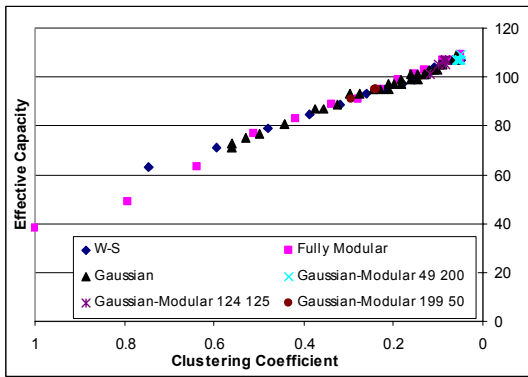
.

Figure 14. Effective Capacity against Clustering Coefficient. $N = 5000$, k=249. Results from four different types of networks are plotted together. A clear linear relationship can be seen in this figure.

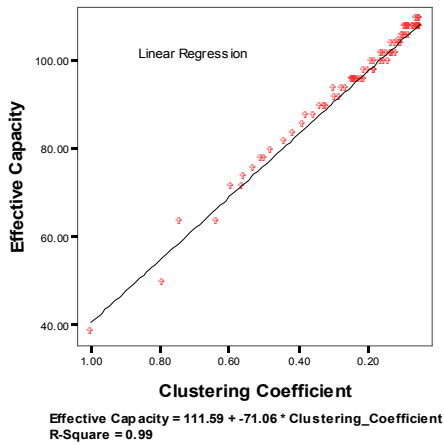Effective Capacity = 111.59 + -71.06 * Clustering_Coefficient
R-Square = 0.99

Figure 15. Linear fit to the data from Figure 14.

A similar analysis was then done for Effective Capacity against Local Efficiency and the results can be seen in Figures 16 and 17. Here *R-Square* is 0.97, so again the fit is highly linear.
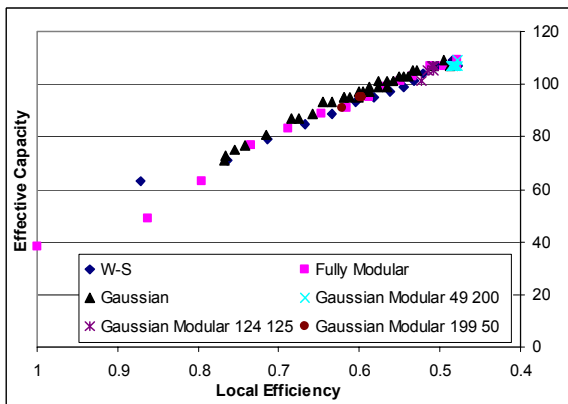

Figure 16. Effective Capacity against Local Efficiency. $N = 5000$, k=249. Results from four different types of networks are plotted together. A clear linear relationship can be seen in this figure.

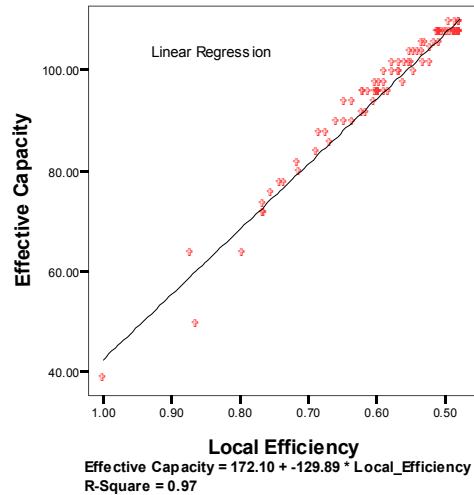Effective Capacity = 172.10 + -129.89 * Local_Efficiency
R-Square = 0.97

Figure 17. Linear fit to the data from Figure16.

The results for networks with $N = 5000$ and $k = 499$ (therefore a network with 10 modules) are very similar (Figure 18, 19). A linear fit is obtained with *R-Square* = 0.99 for both Effective Capacity against Clustering Coefficient and Effective Capacity agianst Local Efficiency. The lines in Figures 15, 17 are different from the ones in Figures 18, 19 because the level of connectivity is different (There are networks with 249 connections per each unit for the first two figures and networks with 499 connections per unit for the second two figures).
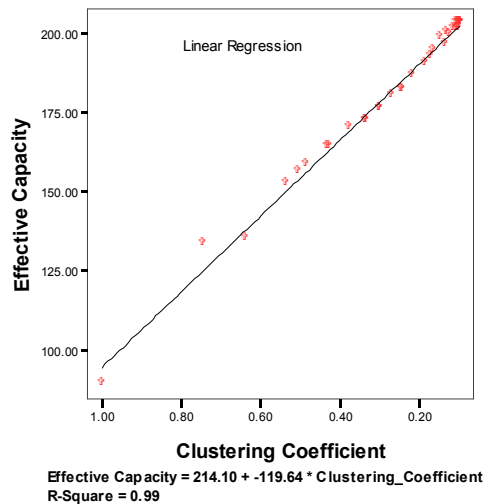

Effective Capacity = 214.10 + -119.64 * Clustering_Coefficient
R-Square = 0.99

Figure 18. Effective Capacity against Clustering Coefficient. $N = 5000$, k=499. Results from four different types of networks are plotted together. A clear linear relationship can be seen in this figure.

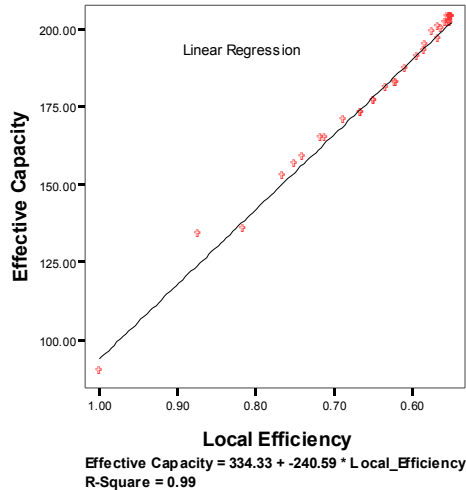*2008 International Joint Conference on Neural Networks (IJCNN 2008)*

Figure 19. Effective Capacity against Local Efficiency. N = 5000, k=499. Results from four different types of networks are plotted together. A clear linear relationship can be seen in this figure.

VII. CONCLUSION

In this paper we investigated how different connectivities affect the performance of high capacity associative memory models. Four different types of networks were examined: a Watt-Strogatz Small-World network, a Modular network, a Gaussian network and a Gaussian Modular network. Several measures of network connectivity were used in the experiments in order to find out any potential correlation in these measures.

Although global features of networks such as Path Length or Global Efficiency are important in determining efficient wiring in the mammalian cortex [4, 10], in the work presented here they show no clear relationship with the associative memory performance. On the other hand, the local clustering (measured by Clustering Coefficient and Local Efficiency) is here shown to have a strong linear relationship with the associative memory performance.

As shown in this paper, this linear relationship seems identical for different types of network models and connectivity distributions. Of course a different connectivity level produces different relationship. This result is potentially important since a purely static measure of network connectivity (clustering) appears to determine an important dynamic property (pattern correction) of the network. One may wonder if this relationship may also govern the associative memory performance in a real mammalian cerebral cortex. Therefore we are currently studying the effect of network connectivity on performance of more biologically plausible models such as spiking neural networks.

REFERENCES

[1] Newman, M.E.J., *Models of the Small World.* Journal of Statistical Physics, 2000. **101**(3 - 4): p. 819.

[2] Watts, D.J. and S.H. Strogatz, *Collective dynamics of'small-world'networks.* Nature, 1998. **393**(6684): p. 409-10.

[3] Calcraft, L., R. Adams, and N. Davey, *Locally-Connected and Small-World Associative Memories in Large Networks.* Neural Information Processing - Letters and Reviews, 2006. **10**(2): p. pp 19-26.

[4] Latora, V. and M. Marchiori, *Economic small-world behavior in weighted networks.* The European Physical Journal B-Condensed Matter, 2003. **32**(2): p. 249-263.

[5] Sporns, O. and J.D. Zwi, *The small world of the cerebral cortex.* Neuroinformatics, 2004. **2**(2): p. 145-62.

[6] Chen, W., et al. *The Connectivity and Performance of Sparce Associative Memory Models.* Progress in Neural Processing - Vol. 17, 2008..

[7] Calcraft, L., R. Adams, and N. Davey, *Efficient architectures for sparsely-connected high capacity associative memory models.* To appear in Connection Science, 2007. **19**(2).

[8] Milgram, S., *The Small World Problem.* Psychology Today, 1967: p. 60-67.

[9] Braitenberg, V. and A. Schüz, *Cortex: Statistics and Geometry of Neuronal Connectivity.* 1998, Berlin: Springer-Verlag.

[10] Watts, D. and S. Strogatz, *Collective Dynamics of 'small-world' networks.* Nature, 1998. **393**: p. 440-442.

[11] Mountcastle, V.B., *The columnar organization of the neocortex.* Brain, 1997. **120**(4): p. 701-722.

[12] Horton, J. and D. Adams, *The cortical column: a structure without a function.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2005. **360**(1456): p. 837.

[13] Abbott, L.F., *Learning in neural network memories.* Network: Computational Neural Systems, 1990. **1**: p. 105-122.

[14] Gardner, E., *The space of interactions in neural network models.* Journal of Physics A, 1988. **21**: p. 257-270.

[15] Calcraft, L., *Measuring the Performance of Associative Memories*, in *Computer Science Technical Report*. 2005, University of Hertfordshire.

[16] de Nooy, W., A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*, in *Structural Analysis in the Social Sciences*. 2005, Cambridge University Press.

[17] Hellwig, B., *A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex.* Biological Cybernetics, 2000. **82**(2): p. 111.