ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

Using graph theoretic measures to predict the performance of associative memory models

Lee Calcraft, Rod Adams, Weiliang Chen and Neil Davey

School of Computer Science, University of Hertfordshire
College Lane, Hatfield, Hertfordshire, AL10 9AB, U.K.
{L. Calcraft, R.G. Adams, W.3.Chen, N. Davey} @ herts.ac.uk

Abstract. We test a selection of associative memory models built with different connection strategies, exploring the relationship between the structural properties of each network and its pattern-completion performance. It is found that the Local Efficiency of the network can be used to predict pattern completion performance for associative memory models built with a range of different connection strategies. This relationship is maintained as the networks are scaled up in size, but breaks down under conditions of very sparse connectivity.

1 Introduction

The seminal paper by Watts and Strogatz [1] on the small-world behaviour of sparsely-connected networks inspired work in a wide range of fields, including the study of neural networks [2-4]. Essential to their argument were the two graph-theoretic measures of *Clustering Coefficient* and *Characteristic Path Length*. A network with local-only connections would have a high Clustering Coefficient, and a long Characteristic Path Length, whereas a randomly-connected network would have a very low Clustering Coefficient.

They argued that networks in nature achieved a compromise between these two parameters, having relatively high Clustering Coefficients, while at the same time relatively short Characteristic Path Lengths. In order to study the relationship between the two measures, they took a locally-connected network and randomly rewired a number of randomly-chosen connections to randomly-selected sites within the network. They found that after a very small amount of rewiring their networks took on the sought-after properties of relatively high Clustering Coefficients and relatively short Characteristic Path Lengths. They named such networks *Small-World* networks.

In 2001 Bohland and Minai [4] applied this technique to a one-dimensional sparsely-connected associative memory model. They found that as the degree of rewiring was increased, the performance of the model improved continuously until the degree of rewiring reached around 40%. By this point the performance of the network had almost reached the level of a random network, and further rewiring had little effect on performance.

Our goal in the present paper is to evaluate to what extent certain graph-theoretic measures can be used to predict this behaviour. To this end we examine the two measures originally used by Watts and Strogatz, the Clustering Coefficient and Characteristic Path Length, together with a new measure named *Local Efficiency*, introduced by Latora and Marchiori [5]. These measures are applied to the underlying graphs of associative memory models built with a range of different connection strategies. It is found that the pattern-completion performance of our models is strongly correlated to the Local Efficiency of the networks from which they are built, to the extent that by measuring the Local Efficiency of a network we can accurately predict pattern-completion performance for a broad class of connection strategies. We will begin by defining the three measures used, and by describing our associative memory model, and the way in which we measure its performance.

2 Characterising sparse directed graphs

The connectivity pattern in an associative memory model may be defined by a connectivity (or adjacency) matrix, $C = \{c_{ij}\}$ where $c_{ij} = 1$ whenever node *i* has an incoming connection from node *j*, and 0 otherwise. The structural properties of graphs can be quantified in terms of the path lengths and clustering of the network.

2.1 Path Lengths

The shortest path length, d_{ij}, between any two nodes in a graph is the minimal number of arc traversals needed to get from one node to the other. The *characteristic path length* is then defined as the mean of these distances: $L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$ where N is the number of nodes in the graph. A problem can arise with this definition if the graph is disconnected as some of the distances will be undefined. For this, and other reasons, Latora and Marchiori [5] introduced the idea of measuring the *global efficiency* of a graph: $E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$ where $\frac{1}{d_{ij}}$ is taken as 0 whenever *i* is not connected to *j*. Note that high path lengths will give low efficiency and vice versa.

2.2 Clustering

Another important measure is the degree to which connections in the graph cluster together. In a social network this is the likelihood of two of your friends also being friends. Watts and Strogatz [1] formalised this with the *Clustering Coefficient*. To calculate this the subgraph G_i is defined as the subgraph made up by the immediate neighbours of node *i* (not including *i*). Then the *Local Clustering Coefficient* C_i is defined as the ratio of the number of edges in G_i to the maximum number of possible edges that could be in G_i. The *Clustering Coefficient* of the complete graph is then defined to be the mean of the C_i. Once again Latora and Marchiori propose a generalisation of this measure that takes into account the distances in G_i. They define the local efficiency of node *i* to be the efficiency of G_i and the Local Efficiency of the graph to be the mean of these individual efficiencies: $E_i = \frac{1}{|G_i|(|G_i|-1)} \sum_{j \neq k \in G_i} \frac{1}{d_{kj}}$ $E_{loc} = \frac{1}{N} \sum E_i$. Here high clustering will imply high Local Efficiency.

3 Network dynamics, training and performance measurement

Our associative memory models consist of a network of perceptrons arranged in a one-dimensional structure with wrap-around at the ends, and the network is trained on sets of random patterns of length *N*, where *N* is the number of nodes in the network. The output of each node is connected to the inputs of a fixed number, *k,* of other nodes. The networks used in the present studies have no symmetric connection requirement [6], and the recall process uses asynchronous random order updates, in which the local field of unit *i* is given by:

$$h_i = \sum_{j \neq i} w_{ij} S_j$$

where w_{ij} is the weight on the connection from unit *j* to unit *i,* and $S\ (=\pm 1)$ is the current state. The dynamics of the network is given by the standard update: $S_i' = \Theta(h_i)$, where Θ is the Heaviside function. Network training is based on the perceptron training rule [7] chosen for its higher resultant capacity than that of the standard Hopfield model. Further details may be found in [8, 9].

Network performance is determined by measuring Effective Capacity [10, 11]. This is a measure of the number of patterns which a network can restore under a specific set of conditions. The network is first trained on a set of random patterns. Once training is complete, the patterns are each randomly degraded with 60% noise, before presenting them to the network. After convergence, a calculation is made of the degree of overlap between the output of the network, and the original learned pattern. The Effective Capacity of the network is the highest pattern loading at which this mean overlap for the pattern set is 95% or greater. The Effective Capacity of a network has been shown to track its underlying maximum theoretical capacity for fully-connected networks [10].

4 Results and Discussion

In the first experiment we took a 500-node one-dimensional network with periodic boundary conditions, and connected it locally so that each node was connected to 50 of its nearest neighbours around a ring. We then measured its Effective Capacity, Clustering Coefficient, Characteristic Path Length, and Local Efficiency as the network was progressively rewired in steps of 10% up to a full 100%, following the technique introduced by Watts and Strogatz [1]. The results appear as Fig. 1, with the Effective Capacity scaled (by dividing it by 20) to fit it on the same graph.

As first demonstrated by Bohland and Minai [4], the pattern completion performance of the network (as measured in this case by Effective Capacity) increases with rewiring up to the point where the rewiring reaches around 40%, after which, little further improvement is achieved. In comparing this behaviour with the structural properties of the underlying graph, we see immediately that Characteristic Path Length appears to be a poor indicator of performance in that it drops from unity to a value of just 0.2 as the local network is rewired by just 1%, whereas associative memory performance (as measured by Effective Capacity) barely increases at all. The other two measures, however, Clustering Coefficient and Local Efficiency, both vary approximately as the inverse of the Effective Capacity.

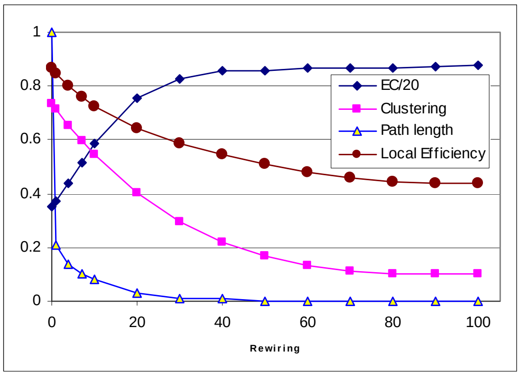
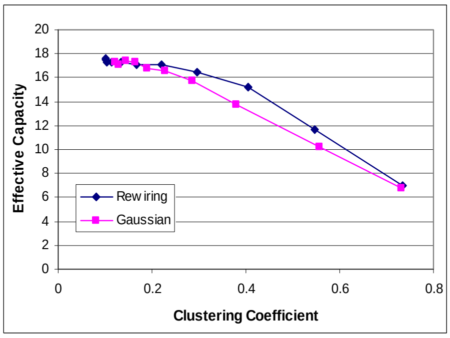


Fig. 1: Effective Capacity, Clustering Coefficient, Characteristic Path Length and Local Efficiency *vs* degree of rewiring for a network of 500 units, with 50 afferent connections per node. Results are averages over 50 runs.

In order to assess to what extent the Clustering Coefficient and Local Efficiency might be used as a predictor of performance, a network of the same size, but using patterns of connectivity based on a Gaussian distribution was created, where the probability of a connection between any two nodes was a Gaussian function of the distance between them. We then made measurements of Effective Capacity, Clustering Coefficient and Local Efficiency for varying values of Gaussian σ, starting with a very tight (almost locally connected) distribution, and progressively increasing σ until a very broad distribution was achieved.

Figure 2a shows a plot of Effective Capacity *vs* Clustering Coefficient for the two networks (progressively rewired and Gaussian), while Figure 2b shows a plot of Effective Capacity *vs* Local Efficiency for the same networks.



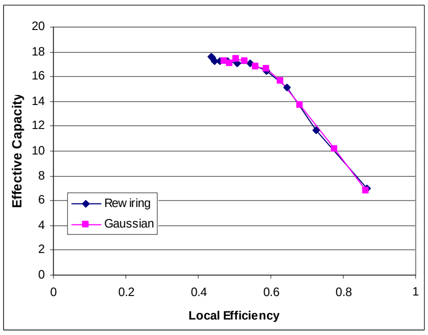


Fig. 2: (a) Effective Capacity *vs* Clustering Coefficient for a network with 500 nodes and 50 afferent connections per node, with patterns of connectivity based on progressive rewiring strategy and Gaussian distributions. (b), as (a) but with Effective Capacity plotted against Local Efficiency. Results are averages over 50 runs in each case.

As may be seen from Figure 2a, the Effective Capacity and Clustering Coefficient of the two networks only coincide at the two extremes of distribution - where the distributions of connections are extremely tight or extremely broad (corresponding to local connectivity or to a random graph). In the case of Figure 2b, however, there is an extremely strong correlation between the Effective Capacity *vs* Local Efficiency plot for both connection strategies. And indeed we have repeated this experiment with different patterns of connectivity - including ones based on exponential distributions, and on restricted uniform distributions, and their plot is inextricable from the curve in Figure 2b

4.1 Larger and more sparse networks

Further experiments were carried out to see if the relationship between Effective Capacity and Local Efficiency maintained for larger and for more sparse networks, and the results appear in Figure 3. Figure 3a is for a network of 2000 units, each with 200 afferent connections. Clearly the relationship still maintains at this larger network size, and interestingly, the Effective Capacity *vs* Local Efficiency curve now approaches linearity.

When we decreased the connection density of the network from 0.1 to 0.001, as in Figure 3b, which is for a network of 5000 units, each with 50 afferent connections, the relationship was no longer maintained, however, with the Gaussian network achieving a higher Effective Capacity for a given Local Efficiency than the progressively-rewired network, in the central region of the graph.

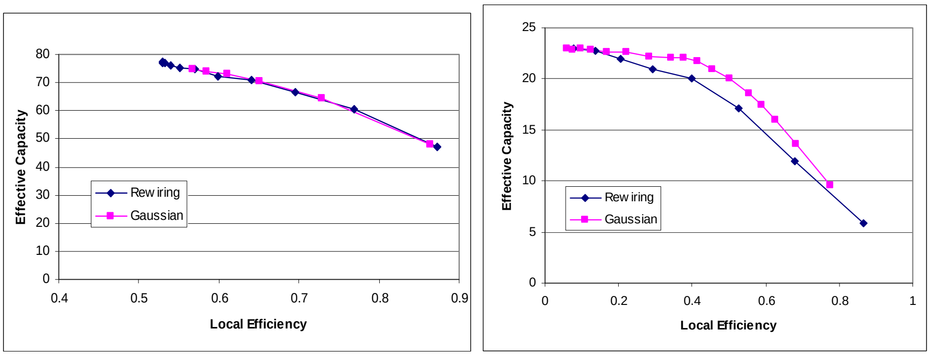


Fig. 3: Effective Capacity vs Local Efficiency for networks based on a progressive rewiring strategy, and on Gaussian distributions. (a) is for a network of 2000 nodes, with 200 afferent connections per node, (b) is for a network of 5000 nodes, with 50 afferent connections per node. Results are averages over 10 runs.

6 Conclusion

In this work we have explored the relationship between the structural properties of different networks, and their pattern-completion performance when used as an associative memory. It was found that of the three graph theoretic measures examined, the Clustering Coefficient, the Characteristic Path Length and the Local Efficiency, one of these, the Local Efficiency, could be used to provide an accurate prediction of pattern-completion performance.

In our first experiments, using a network of 500 units, each with 50 afferent connections, plots of Effective Capacity against Local Efficiency for both progressively-rewired networks, and networks whose pattern of connectivity was based on Gaussian distributions followed precisely the same curve. In other words, by measuring the Local Efficiency of these networks we could predict exactly how many patterns these networks could recall under the test conditions defined by the Effective Capacity measure. This is an important result, especially in view of the dynamic nature of recurrent networks, whose performance is not straightforward to predict mathematically.

These experiments were repeated with a larger network of the same connection density of 0.1 (2000 units with 200 connections), and with a network of connection density of 0.01 (5000 units with 50 connections). It was found that in the case of the former the relationship was maintained, with the Effective Capacity *vs* Local Efficiency curve now approaching a straight line. In the case of the latter more sparse network, of connection density 0.01, the relationship was no longer maintained.

References

[1] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440-442, 1998.

[2] O. Sporns and J. D. Zwi, "The small world of the cerebral cortex," *Neuroinformatics*, vol. 2, pp. 145-162, 2004.

[3] L. G. Morelli, G. Abramson, and M. N. Kuperman, "Associative memory on a small-world neural network," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 38, pp. 495-500, 2004.

[4] J. Bohland and A. Minai, "Efficient associative memory using small-world architecture," *Neurocomputing*, vol. 38-40, pp. 489-496, 2001.

[5] V. Latora and M. Marchiori, "Economic small-world behavior in weighted networks " *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 32, pp. 249-263, 2003.

[6] N. Davey, L. Calcraft, and R. Adams, "High capacity, small-world associative memory models," *Connection Science*, vol. 18, pp. 247-264, 2006.

[7] 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.

[8] N. Davey, S. P. Hunt, and R. G. Adams, "High capacity recurrent associative memories," *Neurocomputing*, vol. 62, pp. 459-491, 2004.

[9] L. Calcraft, R. Adams, and N. Davey, "Efficient architectures for sparsely-connected high capacity associative memory models," *Connection Science*, vol. 19, 2007.

[10] L. Calcraft, "Measuring the performance of associative memories," University of Hertfordshire Technical Report (420) May 2005.

[11] L. Calcraft, R. Adams, and N. Davey, "Locally-connected and small-world associative memories in large networks," *Neural Information Processing - Letters and Reviews*, vol. 10, pp. 19-26, 2006.