Title

Connectivity in Real and Simulated Associative Memories

Authors

Adams, Rod
Calcraft, Lee
Davey, Neil

Connectivity in Real and Simulated Associative Memories

Neil Davey (N.Davey@herts.ac.uk)
Lee Calcraft (L.Calcraft@herts.ac.uk)
Rod Adams (R.G.Adams @herts.ac.uk)
School of Computer Science
University of Hertfordshire
Hatfield, UK

Abstract

Finding efficient patterns of connectivity in sparse associative memories is a difficult problem. It is, however, one that real neuronal networks, such as the mammalian cortex, must have solved. We have investigated computational models of sparsely connected associative memories and found that some patterns of connectivity produce both good performance and efficient use of resources. This could illuminate how real biological systems solve the problem. . Our major finding is that a Gaussian distribution of connection lengths, with a relatively small variance, gives networks with high performance and minimal wiring.

Keywords: Associative Memory; Connectivity; Sparse Networks.

Introduction

In recent years much has been discovered about the pattern of connectivity of the neurons in real neuronal networks. A good deal of this research has explored the way that specific neuronal circuits operate and how functional areas in the brain are connected. Another approach, however, has been to examine the connectivity matrix at the level of individual neurons or of functional areas, and to look for patterns of connectivity familiar in other contexts. This ties in with the large amount of current interest in small-world and scale-free networks. In this paper we look at this second issue in the context of associative memory networks. We summarize some of what is currently known about the connectivity matrix in real neuronal systems and show how the search for optimal patterns of connectivity in artificial associative memory networks may shed light on some of the issues that real neuronal systems must deal with. Our major finding is that a Gaussian distribution of connection lengths, with a relatively small variance, gives networks with high performance and minimal wiring.

Connection Strategies

The connectivity in real neuronal systems, such as the mammalian cortex, is quite different from that found in most artificial neural networks. Real neuronal systems have vast numbers of neurons connected to only a fraction of the other neurons. For example the human cortex has about 10^{11} neurons with each connected to roughly, on average, 10,000 other neurons (Braitenberg & Schüz, 1998).

In such *sparse* networks the connection strategy employed has to balance two competing goals. Firstly the total amount of neuronal fiber should be minimized, both because it is biologically expensive and because increasing length implies increasing difficulty in finding physical paths for the fiber, in the brain. Secondly, however, information needs to travel efficiently throughout the system for fast global computation to take place. The connectivity satisfying both goals is therefore highly likely to be have been optimized by evolution and will be far from random.

Non-Random Graphs

Before looking at some of what is currently known about the actual connectivity strategy employed in real systems, we examine what is known about connection graphs in general.

There is a long history of research into the properties of random graphs, graphs in which the connectivity matrix is randomly configured, often with a specific probability of connectivity (Bollobas, 2001). Recently, however, there has been an explosion of interest in networks with non-random connectivity graphs, such as *small-world* and *scale-free* networks.

The seminal paper of Watts and Strogatz (Watts & Strogatz, 1998) formalised the notion of a *small-world* network. The idea was inspired by work in the Social Sciences showing that there appeared to be only roughly 6 degrees of separation (by acquaintance) between any two people in North America (Milgram, 1967); this despite the fact that most people have a cliquish group of acquaintances, in the sense that any two of their acquaintances are also likely to be acquaintances. The *small-world effect* is therefore characterised as a network with short path lengths (the minimum number of arc traversals to get from one node to another), between any pair of nodes. The simplest sort of network that displays this characteristic is a random network. In a regular random network of N nodes, with each node having k connections, the number of first order acquaintances is k, second order is about k^2, third order k^3 and so on. So in general the number of degrees of separation, D, to reach all N nodes in the network is given by setting $k^D = N$, which gives $D = \dfrac{\ln N}{\ln k}$, so that D increases logarithmically with the size of the network – the small-world effect. However, random

networks are not cliquish and require a relatively large amount of wiring. Watts and Strogatz gave a mechanism for constructing networks that showed the small-world effect, from local networks. Their idea was to begin with a local network and then to *rewire* a small proportion, *p*, of the connections to random targets. Even at very low levels of rewiring, the mean path length between any pair of nodes drops to a value comparable to that of a random network; the rewired connections act as shortcuts through the network. We refer to networks constructed in this way as *small-world networks*.

The cliquishness of a network can be formalised by its *clustering coefficient*, the average fraction of pairs of neighbours of a node, which are also neighbours. Networks that show the small-world effect, but which also have high clustering coefficients have been shown to be remarkably common. Some examples include (Newman, 2000): networks of movie actors, where neighbours are defined by having been in the same movie, power grid networks, the Internet and from our point of view most interestingly, real neuronal networks.

Other interesting networks that show the small-world effect are so called *scale-free* networks (Barabasi *et al.*, 1999; Keller, 2005). These are network models where the distribution of connections follows a power law (that is the frequency of nodes with connectivity *k* falls off as $k^{-\alpha}$). This degree distribution is surprisingly close to that of the distribution of links in the World Wide Web. Some nodes end up with very high levels of connectivity, and act as network hubs, that facilitate short path lengths. Such networks can arise due to a preferential growth process in which nodes that are already well connected are favoured by new connections.

Connectivity in Real Neuronal Networks

The neuronal network of the nematode worm *C. Elegans* has been completely mapped. It consists of 302 neurons and around 1000 connections   A recent analysis (C Cherniak, 1994) of the optimality of the positioning of the neurons (for the given connectivity and physical position of actuators and sensors in the worm) with respect to the total length of wiring (the sum of the length of neuronal fibre) has shown that no better positioning can be found by exhaustive search; a remarkable triumph for evolutionary optimisation. The network also displays short path lengths, an average of 2.65 steps between any two neurons, and a relatively high clustering coefficient of 0.28 (as against 0.05 in an equivalent random network). In (Shefi *et al.*, 2002) cultured in-vitro neuronal networks are studied. They vary in size from *N* = 104 to *N* = 240. Once again the networks show the small-world effect and are relatively highly clustered.

Larger neuronal networks found in more sophisticated animals are not as well understood. Nonetheless several studies have been undertaken into the positioning and connectivity of the neuronal systems. Analysis of the mammalian cortex has been undertaken at two levels of granularity, firstly at the level of the positioning and

connectivity of distinct functional areas such as V1 or V2 in the visual cortex. And secondly at the level of individual neurons. In the first case it has been shown that once again positioning is highly optimised to minimise connection length (Christopher Cherniak *et al.*, 2004; Hilgetag & Kaiser, 2004; Laughlin & Sejnowski, 2003). It has also been shown that the connectivity gives both a small-world effect and a high clustering coefficient (Sporns & Zwi, 2004). The question of whether these neuronal systems show the characteristics of scale-free networks is still open, with opinions differing. (Eguiluz *et al.*, 2005; Sporns & Zwi, 2004).

At the level of individual neurons the connectivity pattern is so complex that only generalised statistics can be produced. These show that in the mouse cortex, for example, there are about 1.6 million neurons, with each connected to, on average, about 8000 other neurons (Braitenberg & Schüz, 1998). The density of connectivity is impressive, with approximately a billion synapses in each cubic millimetre of 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 (Hellwig, 2000), see Figure 1. It is thought extremely unlikely that these intra-area connections are highly structured (Braitenberg & Schüz, 1998) as they are added at the rate of about 40,000 a second as the cortex matures. Cortical connectivity is of particular interest, as it is likely that one major function of the cortex is to act as a very large associative memory (Braitenberg & Schüz, 1998).
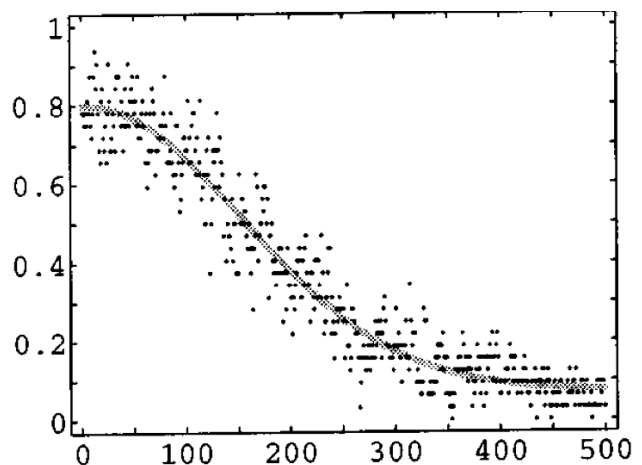

Figure 1: The probability of a connection between any pair of neurons in layer 3 of the rat visual cortex against cell separation. The horizontal axis is marked in μm. Taken with permission from (Hellwig, 2000).

Associative Memory Model

In the work presented here we investigate how the connectivity in a simulated associative memory is affected by the spatial organization of the connections. The model is a variant of the standard Hopfield network. The networks are (for the most part) large and sparse, with 5000 units,

each with incoming connections from 50 other units. Each unit is a simple perceptron, which fires when its net input is greater than zero. The connectivity is not required to be symmetric, and in fact it is unlikely that any two units will be connected in both directions. The networks are presented with training sets of random bipolar (± 1) 5000-ary vectors to learn. The training is done using simple perceptron learning, an iterative procedure guaranteed to converge on a solution if one exists. The network dynamics we employ is random asynchronous updates without replacement. A full description of the technical aspects of the network can be found in (Davey & Adams, 2004).

Such a network can store up to $2k$ patterns (where k is the number of incoming connections that each unit has), and this figure is independent of the specific pattern of connectivity (Davey *et al.*, 2005). However, storing patterns is not the only functional requirement of an associative memory. The other requirement is that the patterns in the training set should be recoverable from noisy versions of themselves: they should be attractors in the state space of the network. As the results presented later show, the pattern of connectivity has a major bearing on this aspect of the network performance.

The metric that we use to measure the performance is the *Effective Capacity* of the network, *EC* (Calcraft, 2005). 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% 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. For large fully connected networks the *EC* value is about 0.1 of the conventional capacity of the network, but for networks with sparse, structured connectivity *EC* is dependent upon the actual connectivity pattern.

Connectivity in the Model

In our model we need the units to have a geometry, so that there is a distance defined between any pair of units. We take the simplest approach (as in the original small-world model (Watts & Strogatz, 1998) and place the units in a 1-D ring. The distance between any two nodes on the ring is simply the minimal number of steps along the ring to get from one to the other.

As already explained there are 2 contrasting patterns of connectivity in such a network: only local connections and random connectivity – see Figure 2.

The local network has minimal total wire length but performs poorly. In our 5000 unit network with 50 connections per unit, the mean connection length is 13. The Effective Capacity is, however, only 5.9 patterns. The explanation for this is that in a local network it is very difficult for a successful global computation to take place. The locality of the connectivity localizes information. The corresponding random network has much more wire, a mean

length 1250, but the performance is much better, an EC of 23 patterns. The question then arises as to whether there is an intermediary distribution of connections that has the performance of a random network, but with significantly reduced wire.
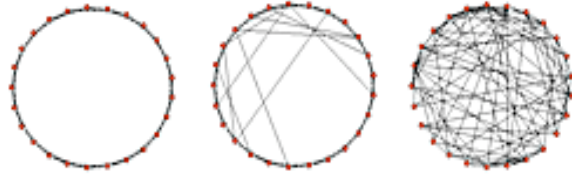

Figure 2: Three types of connectivity, left a locally connected network, center a locally connected network with some random connections and right a randomly connected network.

One obvious contender is a small-world network, as described earlier. Figure 3 shows the frequency of connections at varying lengths for a local network, a small-world network and a random network.
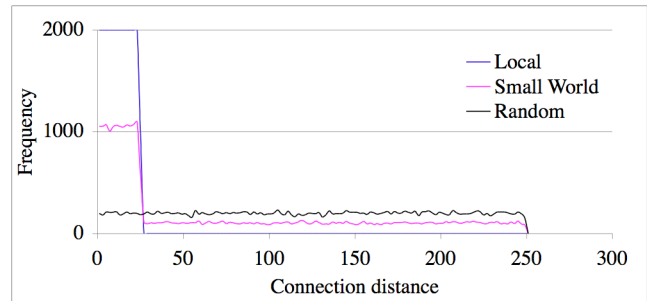

Figure 3: A histogram of the frequency of occurrence of connections at varying lengths in a local network, a small-world network and a random network.

Other patterns are of course possible and two of the most interesting that we have looked at are a Gaussian pattern of fall off and a truncated uniform distribution with a limit to the length of all the connections (L Calcraft *et al.*, 2006). Two examples of such distributions are given in Figure 4.
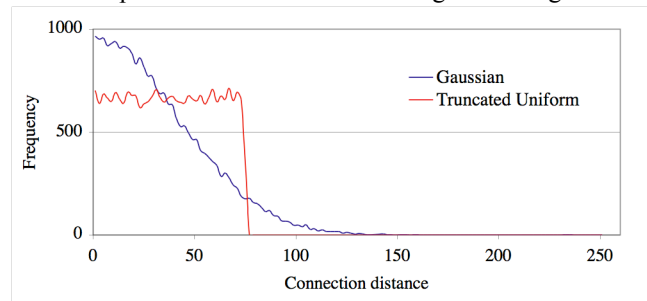

Figure 4: A histogram of the frequency of occurrence of connections at varying lengths in a network with a Gaussian pattern of fall off with standard deviation 40 and a truncated uniform distribution, with length limit of 75.

Results

We have three types of connectivity pattern to evaluate: small-world networks with varying rewiring (which at the two extremes are local and random), Gaussian fall off with varying standard deviations, and truncated uniform distributions with varying maximum allowed connection distance. In order to perform a meaningful comparison we plot the performance of a large number of such networks, with varying characteristics, as shown in Figure 5. The results are rather unexpected. The first observation to be made is that all the network architectures reach optimal performance (an *EC* of about 23) using less wire than the mean length of 1250 needed by the random network. It is also apparent that the small-world networks need much rewiring to achieve this, *p* = 0.6. This still saves wire, as 40% of connections remain local.

The other two distributions, however, perform almost identically and much better than the small-world networks. They both reach optimal performance with mean wire lengths of only 150 per unit. This corresponds to a Gaussian distribution with standard deviation of 130 and a truncated uniform distribution with length limit of 250. The degree of locality of the most parsimonious networks is surprising – these distributions contain no connections that could be described as distal

The almost identical performance of the latter two types of network is remarkable: with the same wiring resource they produce networks with very similar effective capacity.

To shed some light on this Figure 6 gives the actual distributions of connection length when the EC has the value of 20. The major difference between the small-world network and the other two is that it has connections at all lengths, whereas the other two distributions are still fairly local. The conclusion appears to be that these networks can function well without distal connections and show dramatic performance improvement as a completely local constraint is relaxed a little. The detailed pattern of connectivity (Gaussian or truncated uniform) does not appear to be important.

Convergence Times

Nevertheless there may be a hidden cost in the absence of distal connectivity. The time required for the memory to relax to a stable configuration, under the network dynamics, could be significantly increased: information could take longer to move between distant parts of the network. Our early results are interesting and suggest that under low loading all three network types have similar convergence times, but that under stress the truncated uniform distribution falls behind the Gaussian network. Here the presence of some connections of greater length maintains good convergence properties. A fuller analysis of this issue is currently being undertaken.

When considered together these two sets of results suggest that a Gaussian distribution with relatively small variance is a good choice for optimal performance.
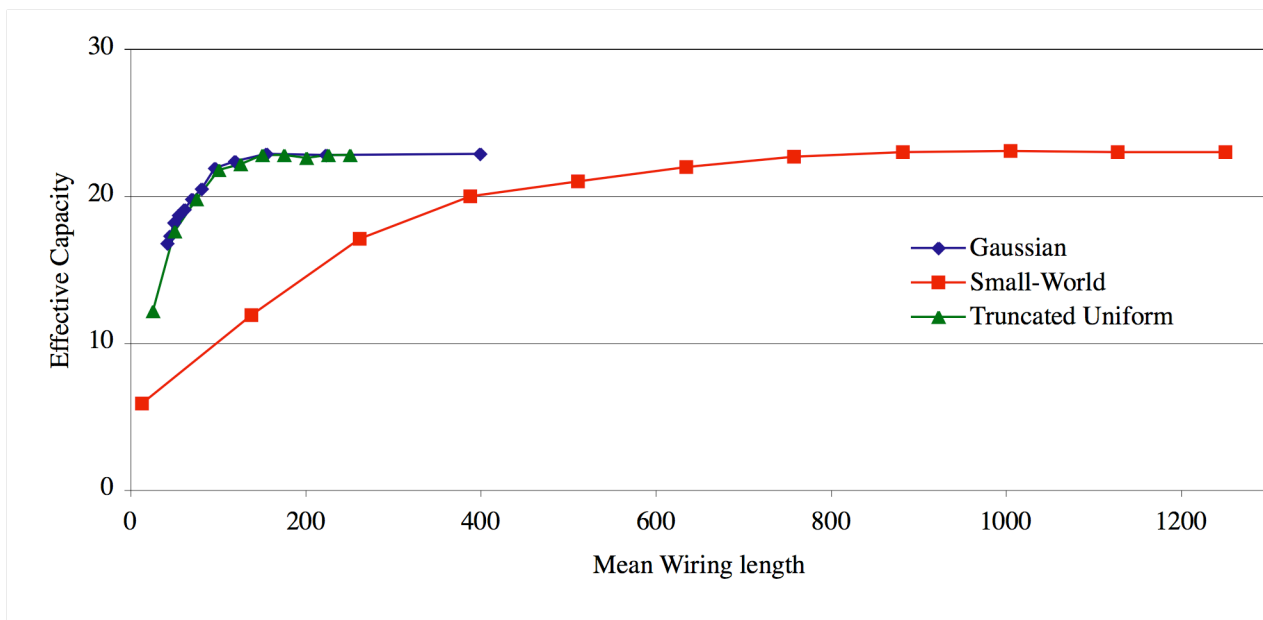

Figure 5: Three different types of network with different variations of wiring. All the networks have 5000 units each with 50 incoming connections. The Gaussian networks have differing standard deviations, from 50 to 500. The Small-World network has rewiring *p* = 0 at the left, and then in increments of 0.1 to a random network *p* = 1 on the right. The Truncated Uniform has varying length restrictions, 50 on the left up to 500 at the rightmost point, in increments of 50. There is a very close fit between the performance of the Gaussian networks and the Truncated Uniform networks, for a given mean wiring length.
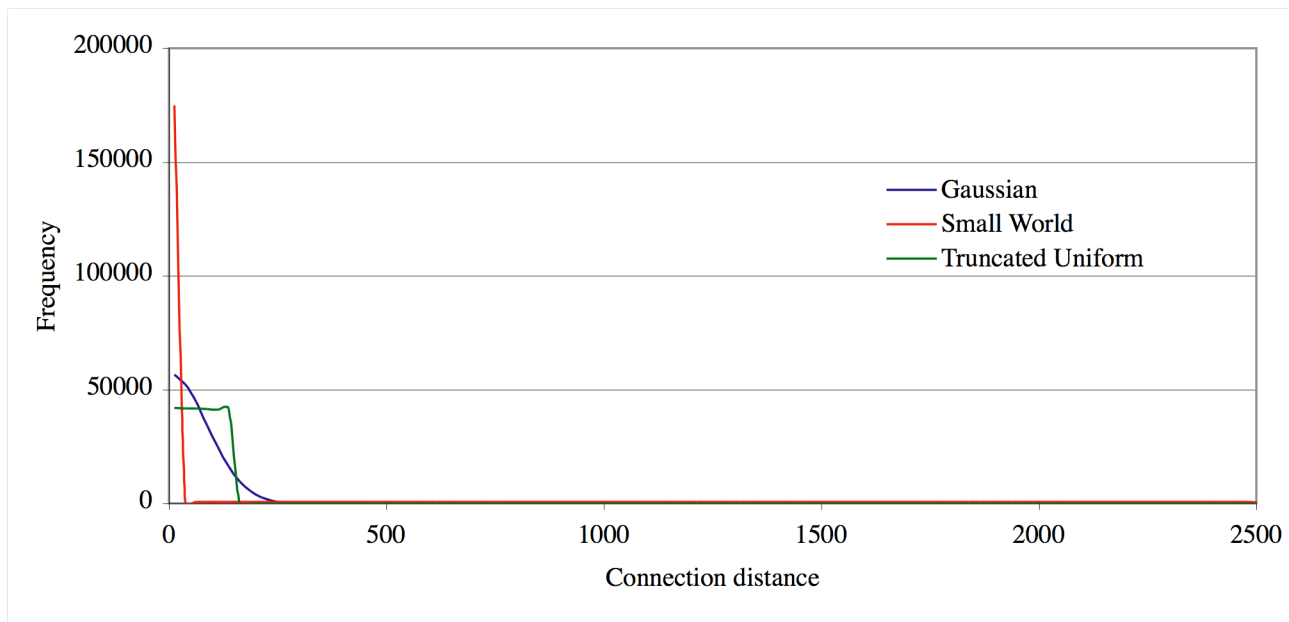
Figure 6: A plot of the distribution of connections in a Gaussian distribution with standard deviation 78, a small-world network with $p = 0.4$ and a Truncated Uniform distribution with maximum connection length of 150. All the networks have 5000 units each with 50 incoming connections and an Effective Capacity of 20.

Evolved Networks

The final investigation we have undertaken takes a quite different approach to finding optimal patterns of connectivity. Rather than imposing a specific distribution we use a genetic algorithm (GA) to find a configuration that maximizes performance whilst minimizing the mean wiring length. Due to the computational complexity of this task we are not able to use networks of the size used earlier. The maximum size network we have been able to evolve is 400 units, with 20 connections per unit. The connection matrix that constitutes the genome is therefore 160,000 bits. The population size was 50. The fitness function plateaus after about 60,000 generations, a runtime of several months. Full details of the algorithm, the representation used and other similar results can be found in (Adams *et al.*, 2005).

The fittest network found by the GA has a mean connection length of about one half of an equivalent random network, but has pattern correction capability no worse than a random network. The final pattern of connectivity found is shown in Figure 7. The distribution is somewhat similar to a Gaussian distribution. However caution should be taken in any direct comparisons between the evolved network and those presented earlier, with specific distributions of connection length. This is because of the order of magnitude difference in the size of the networks and the difference in sparseness.

Discussion

In a computational device such as the mammalian cortex, information must pass efficiently through the system if it is to be integrated at a global level. This puts huge demands on the connectivity in the system, for it is extremely costly to have many long range connections. In the simple model of associative memory presented here, a preference was clearly demonstrated for a Gaussian like probability distribution in both the designed and evolved networks. Moreover in the large and very sparse networks good performance was obtained with highly (but not completely) local connectivity. This is very interesting and unexpected. An important issue raised by this fact is to determine the relationship between the size of the network, the degree of sparseness and the actual variance required for the most parsimonious network. This is an analysis we have yet to undertake.

Of course the model presented here has several shortcomings as a realistic model of the cortex, and addressing these will be interesting areas of future work. Specifically it is of interest to see if our results are robust to models in which:

• The artificial neurons are placed in 2 dimensional layers
• More realistic integrate and fire units are used
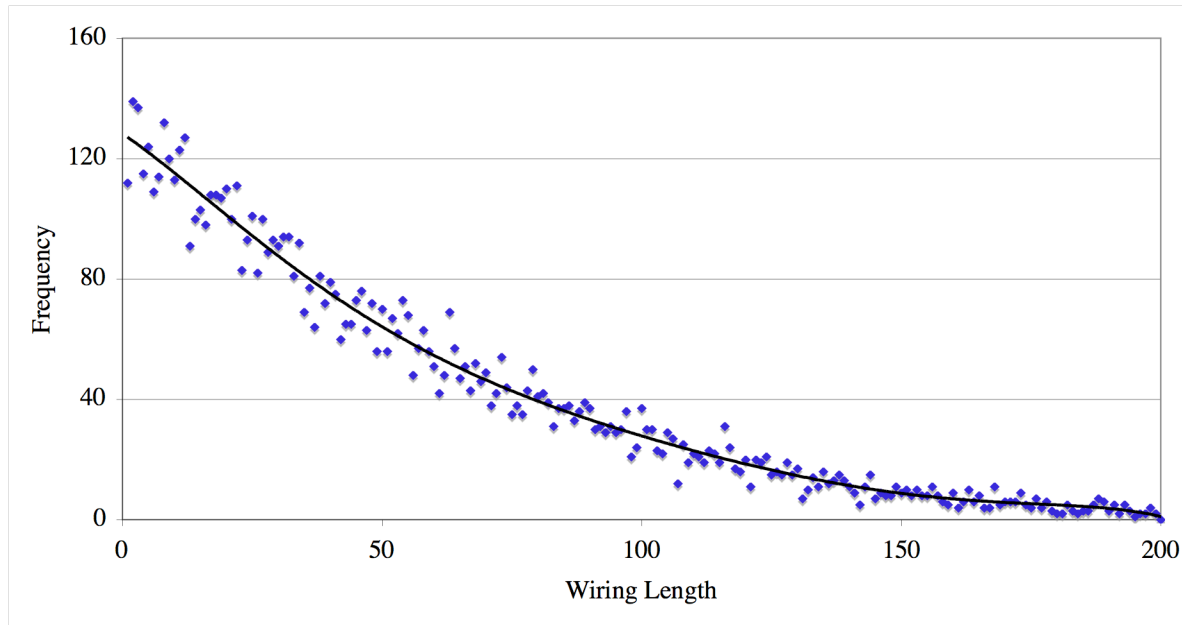• Signal propagation time is taken into account

Figure 7: The pattern of connectivity found by a GA attempting to minimize wiring length and maximize performance. The network has 400 units each with 20 incoming connections. This distribution was found after 60,000 generations.

References

Adams, R., Calcraft, L., & Davey, N. (2005, July). *Evolving high capacity associative memories with efficient wiring*. Paper presented at the IEEE International Joint Conference on Neural Networks, Montreal.

Barabasi, A., Albert, R., & Leong, H. (1999). Scale-free characteristics of random networks: The topology of the world wide web. *Physica A: Statistical Mechanics and its Applications, 272*, 173-187.

Bollobas, B. (2001). *Random graphs*: Cambridge University Press.

Braitenberg, V., & Schüz, A. (1998). *Cortex: Statistics and geometry of neuronal connectivity*. Berlin: Springer-Verlag.

Calcraft, L. (2005). *Measuring the performance of associative memories* (Technical Report No. 420): University of Hertfordshire.

Calcraft, L., Adams, R., & Davey, N. (2006). Gaussian and exponential architectures in small world associative memories, *ESANN*. Bruge.

Cherniak, C. (1994). Component placement optimization in the brain. *J. Neurosci., 14*(4), 2418-2427.

Cherniak, C., Mokhtarzada, Z., Rodriguez-Esteban, R., & Changizi, K. (2004). Global optimization of cerebral cortex layout. *PNAS, 101*(4), 1081-1086.

Davey, N., & Adams, R. (2004). High capacity associative memories and connection constraints. *Connection Science, 16*(1), 47-66.

Davey, N., Calcraft, L., & Adams, R. (2005). *Associative memories with small world connectivity*. Paper presented at the ICANNGA, Coimbra.

Eguiluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., & Apkarian, A. V. (2005). Scale-free brain functional networks. *Physical Review Letters, 94*, 018102.

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

Hilgetag, C., & Kaiser, M. (2004). Clustered organization of cortical connectivity. *Neuroinformatics, 2*(3), 353-360.

Keller, E. F. (2005). Revisiting "scale-free" networks. *BioEssays, 27*(10), 1060-1068.

Laughlin, S. B., & Sejnowski, T. J. (2003). Communication in neuronal networks. *Science, 301*(5641), 1870-1874.

Milgram, S. (1967). The small world problem. *Psychology Today, 2*, 60-67.

Newman, M. E. J. (2000). Models of the small world. *Journal of Statistical Physics, 101*(3/4), 819-841.

Shefi, O., Golding, I., Segev, R., Ben-Jacob, E., & Ayali2, A. (2002). Morphological characterization of in vitro neuronal networks. *Physical Review E, 66*(021905), 021905.

Sporns, O., & Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics, 2*(2), 145-162.

Watts, D., & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440-442.