Multiple Scales in Small-World Networks

Rajesh Kasturirangan
Department of Brain and Cognitive Sciences
MIT
MIT building NE-20, Cambridge, MA 02139
kasturi@ai.mit.edu

Abstract

Small-world architectures may be implicated in a range of phenomena from networks of neurons in the cerebral cortex to social networks and propogation of viruses [1]-[2]. Small-world networks are interpolations of regular and random networks that retain the advantages of both regular and random networks by being highly clustered like regular networks and having small average path length between nodes, like random networks. While most of the recent attention on small-world networks has focussed on the effect of introducing disorder/randomness into a regular network, we show that that the fundamental mechanism behind the small-world phenomenon is not disorder/randomness, but the presence of connections of many different length scales. Consequently, in order to explain the small-world phenomenon, we introduce the concept of multiple scale networks and then state the multiple length scale hypothesis [3]. We show that small-world behavior in randomly rewired networks is a consequence of features common to all multiple scale networks. To support the multiple length scale hypothesis, novel network architectures are introduced that need not be a result of random rewiring of a regular network. In each case it is shown that whenever the network exhibits small-world behavior, it also has connections of diverse length scales. We also show that the distribution of the length scales of the new connections is significantly more important than whether the new connections are long range, medium range or short range.

1 Introduction

A small-world network [1]-[2] is a network obtained by randomly rewiring a small percentage of the connections of a regular network. The role of the regular connections is similar to that of short range, topographic connections in a biological network while the rewired random connections play the role of medium and long range connections in a biological network. Small world networks have been shown to have very interesting computational properties [1]-[2]. For example,the average path length between nodes is much smaller in small-world networks than in regular networks, even though the local topology of small-world networks is more or less the same as that of regular networks.Also, coupled oscillators on a small-world network converge very quickly[1]-[2]. Furthermore, small-world networks seem to be quite reasonable models for various biological networks. For example,the neural network of *c.elegans* fits the small-world framework[1]-[2].

There are strong computational reasons for expecting small-world type architectures in biological neural networks. Fast convergence of oscillator nets is likely to be important in quickly computing scene interpretations. Also, as Marr pointed out [4], it is unlikely that the brain has time for more than a few iterations when trying to converge to a percept. Consequently, the architecture of the brain should be of a kind that allows convergence with only a few iterations (which is a defining property of small-world networks).

However, randomly rewired networks are not realistic models of real world biological networks. This is because the medium-range and long-range connections in a biological network are unlikely to be random. On the other hand,it is quite likely that the distribution of connections in a network plays an important role in determining the computational properties of the network. One might hypothesize that the distribution of short range connections versus medium range and long range connections is responsible for various observed properties of biological networks, which is a biologically plausible hypothesis. For example, it has been shown that in area V1, there are both short range and long range connections [5]. The above discussion raises some interesting questions, such as-

(1) How essential is randomness to the small-world phenomenon? Are there other network architectures with strong clustering and with low average path length?
(2) What is the role of short range, medium range and long-range connections in determining the computational and dynamical properties of a biological network?

We show that there is a simple answer to questions (1) and (2), namely, multiple scale networks and the multiple length scale hypothesis respectively. In this paper, we first define multiple scale networks followed by the statement of the multiple length scale hypothesis. Three different network architectures are introduced. Numerical simulations of these architectures lend support to the multiple length scale hypothesis. Then small-world behavior for randomly rewired networks is derived as a consequence of properties common to all multiple scale networks.

2 Multiple scale networks and the multiple length scale hypothesis

2.1 Multiple scale networks

Definition 1. Let N be a network with n nodes. Let the distance between two nodes, f and g in N be denoted by $\text{dist}_N(f,g)$. Let w new connections be added to N, forming a new network N'. Call the set of new connections W, so that cardinality(W) = w. For each new connection e_w, $\text{dist}_N(e_w)$ is the distance- $\text{dist}_N(f_e, g_e)$ - in N between the endpoints f_e and g_e of e_w. The distribution of length scales in W is measured by the function-

$$D:\ W \mapsto (1,\infty)\ ,\ D(e) = \text{dist}_N(e). \tag{1}$$

Definition 2. A network N' obtained by adding connections to the network N is multiple scale with respect to N (denoted $N' \preceq N$) if -

$$\begin{array}{l} \exists r \gg 0,\ \text{ and } \exists \text{ length scales } l_i, i = 1,2\ldots r \text{ such that} \\ 0 < l_1 \ll l_2 \ldots \ll l_r\ \leq n \text{ and} \\ \forall i : i \leq r, l_i \in D(W). \end{array} \tag{2}$$

Definition 3. A network N is a **multiple scale network** if it has a subnetwork M with the same number of nodes such that $N \preceq M$. Note that M contains all the nodes of N but not all the connections of N.

Suppose we start with a regular network $R_{n,k}$ with n nodes and k connections per vertex and add w new connections to $R_{n,k}$ forming a new network $R_{n,k,w}$. Furthermore, if $w \gg$ log$n$ and the distribution of the new connections is uniform and random-

$$R_{n,k,w} \preceq R_{n,k} \text{ and } R_{n,k} \subseteq R_{n,k,w}. \tag{3}$$

Therefore $R_{n,k,w}$ is a multiple scale network. Equation (3) holds because the new connections are chosen at random,and as a consequence, cardinality$(D(W)) \gg w \gg$ log$n$. Is it possible that the randomly rewired network $R_{n,k,w}$ is a small-world network (as shown in Watts and Strogatz [1]) *because* it is a multiple scale network?

2.2 The Multiple length scale hypothesis

The reduction in average path length when rewiring a network is proportional to the number of length scales present in the new connections and the number of connections at each length scale. In other words, *any* network, random or deterministic, that is a result of adding sufficiently many connections at many different length scales to a regular network, will exhibit small-world behavior. Moreover, the small-world behavior is not dependent on whether the new connections are long range or short range. What matters is the *number of scales* represented in the new connections.

3 Network architectures

We now review three different network architectures, the first an analytically tractable example, the second related to communication networks and the third to social networks. In each case, it is shown that the numerical data supports the multiple length scale hypothesis.

3.1 A regular network coupled with a tree structure (R+T networks)

For the sake of simplicity, assume that $n = k2^m$. Then, we can divide N into 2^m blocks, each of size k. At this point, connections are drawn from the midpoint of each block to all the other points in the block, so the R+T network has a regular subnetwork $R_{n,k}$. Take all the midpoints m_j and number them from 0 to $2^m - 1$. The tree structure is defined on the set of midpoints of the regular subnetwork in the following manner-

For each pair of midpoints m_i and m_j, there is a an connection $e_{i,j}$ connecting m_i and m_j if and only if $i - j$ is a power of 2, i.e., $\exists\, r \le m - 1$ such that $i - j = 2^r$ (4)

In an R+T network, the total number of connections is $nk/2 + 2^m = k^2 2^{m-1} + 2^m$ which is- $k/2 + 1/k$ - connections per vertex .Therefore, the tree structure adds only $1/k$ connections per vertex. Moreover, it is not hard to show that the average path length, $L_{\text{R+T}} = m + 1 \cong \log_2 m/k$. Therefore, in an R+T network with $n = 256$, $k = 8$ and $m = 5$, the number of connections added by the tree structure is 32 and $L_{\text{R+T}} = 6$. In comparison, in a computer simulation, starting with the same regular network, 32 randomly chosen connections were added. We got $L_{\text{random}} = 7.1 > L_{\text{R+T}}$. Note that the tree structure forces the R+T network to be a multiple scale network, as $R + T \preceq R_{n,k}$. Moreover, the R+T network is quite clustered as we are adding only $1/k$ more connections per vertex.

3.2 Hub creation and the maximum rule

Suppose we want to reduce the average path length in a communications network, where the wiring diagram of the network is known. In this case, what we may want to do is to minimize path length while using as few connections as possible. One way of doing this is

Table 1: Comparison of Maximum rule and Random rewiring rule										
	Maximum rule					**Random rewiring rule**				
E^{i}	825	825	825	825	825	825	825	825	825	825
w	5	10	25	50	100	5	10	25	50	100
r_{f}	3	5	9	15	18	5	8	13	19	20
L^{i}	25.14	24.63	25.00	24.25	24.83	24.81	25.16	24.90	25.42	25.14
L^{f}	18.05	14.12	9.60	7.12	5.42	16.42	12.73	9.97	8.11	5.97

to create *hubs* , i.e., nodes that are connected to many other nodes. In the extreme case, we can create a "universal" hub by adding $n-1$ new connections, from vertex 1 to vertex i for all nodes i . If each vertex in the original regular network has $k \gg \log n$ connections then we are adding only one connection per vertex on average. Moreover, each hub also gives rise to connections of different length scales. However, this algorithm may not be feasible in general. For example, if this is a model for air transport, the above network model leads to overcrowding at the hubs (as has been noticed by everybody who flies out of a major hub in the U.S).

Another solution to the rewiring problem in a communications network is the **maximum rule** . Let $N_t,\ t = 0, 1, 2\ldots$ be a network that is evolving in time. At each time step t, a new connection is drawn between the two points that are most distant from each other in N_{t-1}. In the randomized version of the maximum rule, m pairs of points are selected randomly at each time step t and an connection is drawn between the most distant of the m pairs (in N_{t-1}). The results of a computer simulation of the maximum rule are displayed in table 1 along with the results for a randomly rewired network. In the randomly rewired network, the rewiring is done assuming a uniform distribution of the new connections.

One might think that the maximum rule is the best way to reduce the average path length, given that the maximum rule is explicitly designed to bring the furthest nodes close to each other. However, Table 1 shows that for w = 5, 10 and 25, the random rewiring rule does *as well or better* than the maximum rule which is quite surprising . However, a comparison of r_{random} and $r_{\text{max-rule}}$ explains the counterintuitive behavior. For $w = 5,\ 10$ and 25 we got $r_{\text{max-rule}} = 3,\ 5$ and 9 respectively while $r_{\text{random}} = 5,\ 8$ and 13 respectively. Therefore the number of length scales in the randomly rewired network is more than the number of length scales in the maximum rule network if the number of new connections is the same.

The comparison of the maximum rule and the random rewiring rule suggests that the strongest determinant of small-world behavior is the presence or absence of a variety of length scales (and not whether the new connections are short range or long range). This hypothesis was tested in a computer simulation by selectively removing length scales from a randomly rewired network.. The initial network was a regular network with $n = 275$ and $k = 6$, randomly rewired by adding 50 new connections. Then, (approximately) half of the connections were removed in three different ways- The shortest 1/2 of the connections, the middle 1/2 of the connections and the longest 1/2 of the connections. The average shortest path length was computed for each of the three networks.The results are shown in Table 2 below.

Table 2 shows that for $w = 27$ there is no difference in the performance of the three networks, suggesting that the absolute length of the new connections is not as important as their distribution.

Notation. E^{i} is the number of connections in the initial network N^{i}. w is the number of new connections. L^{i} is the average shortest path length in the initial network N^{i}. L^{f} is the average shortest path length in the final network N^{f}. r_{f} is the number of length scales represented by the new connections w with respect to the distance metric in N^{i}.

Table 2: Selective removal of length scales			
Length scales removed	70 - 138	4-70	30-100
n^{i}	825	825	825
w	27	27	28
L^{i}	24.99	24.91	24.81
L^{f}	10.08	10.04	10.01

Table 3: The FOFF rule				
n^{i}	825	825	825	825
w	50	100	200	500
r_{f}	2	4	6	19
L^{i}	24.95	24.76	24.73	24.86
L^{f}	18.98	14.97	10.23	5.48

3.3 A Friend of a friend is a friend (FOFF)

The FOFF rule is a dynamical rule that is a model of social networks . At each time step t, an connection is drawn between two nodes f and g having a common friend h that are not connected by an connection at time $t-1$. Table 3 shows the result of a computer simulation of the FOFF rule. The decrease in shortest path length for the FOFF rule is much slower than the corresponding decrease for randomly rewired networks (see Table 3 below). For w = 50, 100, 200 and 500 we got L_{FOFF} = 18.98, 14.96, 10.23 and 5.48. The multiple length scale hypothesis explains the behavior of the FOFF rule quite readily because r_{f} grows very slowly as a function of w. In fact, for w = 50, 100, 200 and 500 we got r_{f} = 2, 4, 6 and 19 , which is a lot slower than the results for the maximum rule or the random rewiring rule.

4 Deriving small-world behavior from the multiple length scale hypothesis

Notation. N^{i} will always stand for the initial network and N^{f} will always stand for the final network. Let the set of new connections be denoted by W and the cardinality of W = w. Throughout the ensuing discussion, e will always stand for a new connection that has been added to the initial network N^{i}. For a pair of nodes f and g, the distance between f and g in N^{i} is denoted by $\mathrm{dist}_i(f, g)$ while the distance between f and g in N^{f} is denoted by $\mathrm{dist}_f(f, g)$. Similarly, for each new connection e, $\mathrm{dist}_i(e)$ is the distance (in N^{i}) between the endpoints f^e and g^e of e. r is the number of length scales in N^{f}. $D(e)$ is the distibution of the connections in W. Let $R_{n,k}$ be a regular network with n nodes, k connections per vertex. Let $R_{n,k,w}$ be the network that we get by adding w new connections to $R_{n,k}$.

Tight covering : A length scale l covers the network $R_{n,k,g}$ if for each vertex h in $R_{n,k,g}$, there is a new connection e, $\mathrm{dist}_{\mathrm{reg}}(e) \cong l$ such that x lies between the endpoints f and g in the regular network $R_{n,k}$. In other words, the new connections of scale l wrap around the network. Furthermore, in order to reduce average path length, it is not enough that a length scale l cover the network. Suppose there are two nodes f and g such that $\mathrm{dist}_{\mathrm{reg}}(f, g) \cong l$. Ideally, we want the shortest path between f and g in $R_{n,k,w}$ to contain connections of length $\cong l$ only. In order to ensure this property of the shortest path, we have to make sure that the end point of an connection of length $\cong l$ is close to the starting point of another connection of length $\cong l$, i.e., the new connections of length $\cong l$ are tightly packed. In such a situation, we say that a covering of the network by connections of the length scale l is *tight* .

Let $R_{n,k,g}$ be a regular network that has been rewired by adding w new connections. Furthermore assume that there is a scaling factor s and length scales $s^i,\ i \leq \log_s n$ such the new connections are chosen randomly and that their distribution is uniform with respect to the length scales, i.e.,

$$P\left(e \in W\ : s^k \leq \text{dist}_{\text{reg}}(e) \leq s^{k+1}\right)\ =\ P\left(e \in W\ : s^l \leq \text{dist}_{\text{reg}}(e) \leq s^{l+1}\right) \\ \forall\, 0 \leq\ k, l\ \leq\ \log_s n \tag{5}$$

Let $k \cong\ \log_s n$ and let $w \cong\ n\ \ll\ nk$. (5) implies that the number of connections per length scale is approximately $\frac{n}{\log_s n}$. If $s^i\ >\ \log_s n$, the connections of that length scale almost certainly cover $R_{n,k,w}$ (see the beginning of this section for a definition of "cover"). The average distance between successive connections of length s^i is about $\log_s n\ \cong\ k$, which means that in $R_{n,k,w}$ successive connections are separated by about an connection since $k \cong\ \log_s n$. Therefore, if $s^i\ >\ \log_s n$, the connections of scale s^i tightly cover the network. Let f and g be two nodes such that

$$\log_s n\ \leq\ s^i\ \leq\ \text{dist}_{\text{reg}}(f, g)\ \leq\ s^{i+1},\ i \leq\ \log_s n. \tag{6}$$

The following inequality is an easy consequence of the fact that the scales tightly cover the network-

$$\text{dist}_f(f, g)\ \leq\ 2s + \text{dist}_f(h, g) \tag{7}$$

Where h is an intermediate point such that $s^{i-1}\ \leq \text{dist}_{\text{reg}}(h, g)\ \leq\ s^i$. Inequality (7) gives rise to

$$\text{dist}_f(f, g)\ \leq\ 2s.\log_s(\text{dist}_{\text{reg}}(f, g)) \tag{8}$$

showing that the distance in the final network is uniformly logarithmic for all pairs f and g, replicating the qualitative behavior of small-world networks.

Shortest path algorithms. Given two nodes in a network how can we find the shortest path between the two? In a random network there seems to be no way out besides checking all paths and then picking the shortest of the lot. From a computational point of view this is an extremely expensive algorithm, involving $O(n^2)$ computations on average. The same problem seems to arise in the case of a regular network modified by adding a few random connections. Even though the network is small-world and the behavior of the average shortest path length is known, there is no efficient way of determining the shortest path between two points. Once again, the multiple length scale hypothesis helps in formulating a better algorithm for the problem of finding the shortest path between two nodes.
Suppose a network N has many different length scales $s^i,\ i\ \leq\ \log_s n$. Furthermore assume that all the length scales s^i tightly cover the network N. Let f and g be two nodes in N. Let h be an intermediate vertex in the shortest path P(f,g) from f to g. Define

$$k = max\left\{l : s^l\ \leq\ \text{dist}_{\text{reg}}(f, g)\right\} \\ \Delta = \left\{e\ \in\ W\ :\ s^k\ \leq\ \text{dist}_{\text{reg}}(f, g)\ \leq\ s^{k+1}\right\} \\ \Theta\ =\ \{\exists z\ :\ e\ \in\ \Delta \text{ such that } e \text{ is an connection from } h \text{ to } z\} \tag{9}$$

Then, there is an connection connection e_h in P(f,g) starting at h (going towards g) satisfying the following properties-

$$e_h\ \in\ \Delta \text{ and if } y \text{ is the other vertex of } e_h,\ \text{then} \\ \text{dist}_{\text{reg}}(y, g)\ \leq\ \text{dist}_{\text{reg}}(y', g)\quad \forall\, y'\ \in\ \Theta \tag{10}$$

Two properties of the shortest path algorithm follow from (10)
(i) The shortest path algorithm is local, i.e., at each intermediate vertex h in the shortest path P(f,g) from f to g, the next connection in P(f,g) is in the neighborhood of h.
(ii)At each intermediate vertex h, the shortest path algorithm considers a small subset of the connections in the local neighborhood of h, namely the ones that are of the right scale.

From (i) and (ii) it follows that the shortest path algorithm performs $O(\log_s(\text{dist}_{\text{reg}}(f, g)))$ computations, which is $\cong O(\log_s n)$ computations when computing the shortest path from f to g. In comparison, for an arbitrary network we need to perform $O(n^2)$ computations.

5 Conclusion

There is no universal network architecture that is an optimal solution to all problems. Nevertheless it would be extremely useful to have a class of network architectures that are near optimal for a wide variety of problems. Networks with connections of multiple scales are an especially rich class of models that provide a unified framework for describing and modeling a variety of networks [3]. Multiple scale networks can be used to model both random and deterministic networks. This paper shows that properties common to all multiple scale networks explains the small-world phenomenon and that the shortest path problem is also much simpler to solve in multiple scale networks.

Multiple scale networks are tractable, both analytically and by computer simulations. We will be better equipped to understand the behavior of real world networks if we focus our attention on the distribution of length scales in networks as opposed to the effect of introducing increasing amounts of disorder into a regular network.

Methods: A note about the algorithm and the tables

The same network algorithm is used in all the simulations. All the rewiring rules mentioned in this paper are initialized on a regular network with $n = 275$ nodes and $k = 6$ connections per vertex. 1000 pairs of nodes are then selected at random and the shortest distance between all the pairs is found. The algorithm also keeps track of the number of connections added at any stage. All the networks are plots of the existence or non-existence of a length scale in the new connections. For a network with 275 nodes, the value of the function $\text{dist}_{\text{reg}}(f, g)$ is divided into 20 blocks. For each new connection with nodes f and g, $\text{dist}_{\text{reg}}(f, g)$ is computed and then the Y- value of the appropriate block between 1 and 20 is changed from 0 to 1.

Acknowledgements

I want to thank Whitman Richards for many stimulating discussions about multiple scale networks (and about life, the universe and everything in between) and Josh Tenenbaum for many useful comments regarding the style and substance of this paper.

References

[1] Watts, D. and Strogatz, S., The collective dynamics of small-world networks, *Nature* , 1998 June 4;393(6684): 440-442.

[2] Watts, D., *Small Worlds : The Dynamics of Networks between Order and Randomness* , To be published by Princeton University Press.

[3] For more details, see Kasturirangan, R., Multiple Scales in Small-world Networks, http://xxx.lanl.gov/abs/cond-mat/9904055.

[4] Marr, D., *Vision* , W.H.Freeman, 1982.

[5] Gilbert,C.D., and Wiesel,T.N.,Clustered intrinsic connections in cat visual cortex, *Journal of Neuroscience* , 1983,May 3(5): 1116-1133.