Small worlds: How and why

Nisha Mathias[1,*] and Venkatesh Gopal[2,†]
[1]*Department of Computer Science and Automation, Indian Institute of Science, Bangalore 560 012, India*
[2]*Raman Research Institute, Sadashivanagar, Bangalore 560 080, India*
(Received 29 May 2000; published 26 January 2001)

We investigate small-world networks from the point of view of their origin. While the characteristics of small-world networks are now fairly well understood, there is as yet no work on what drives the emergence of such a network architecture. In situations such as neural or transportation networks, where a physical distance between the nodes of the network exists, we study whether the small-world topology arises as a consequence of a tradeoff between maximal connectivity and minimal wiring. Using simulated annealing, we study the properties of a randomly rewired network as the relative tradeoff between wiring and connectivity is varied. When the network seeks to minimize wiring, a regular graph results. At the other extreme, when connectivity is maximized, a ''random'' network is obtained. In the intermediate regime, a small-world network is formed. However, unlike the model of Watts and Strogatz [Nature **393**, 440 (1998)], we find an alternate route to small-world behavior through the formation of hubs, small clusters where one vertex is connected to a large number of neighbors.

DOI: 10.1103/PhysRevE.63.021117 PACS number(s): 05.40.−a

I. INTRODUCTION

Coupled systems can be modeled as networks or graphs, where the vertices represent the elements of the system, and the edges represent the interactions between them. The topology of these networks influences their dynamics. Network topologies may be random, where each node or vertex is randomly wired to any other node; or they may be regular, with each vertex being connected in a fixed pattern to an identical number of its neighboring nodes [1]. Watts and Strogatz [2] showed that between these two extremes of topology lies another regime of connectivity, which they call *small-world* networks. Such networks are ''almost'' regular graphs, but with a few long range connections.

Consider a few examples of networks: neurons in the brain, transportation and social networks, citations of scientific papers, and the World Wide Web. While some of these networks come from the physical world and others do not, nevertheless they all have a well-defined cost metric associated with them. Neural and transportation networks are networks where the cost of a connection translates to the physical distance between its adjacent nodes. In nonphysical networks, such as social networks, paper citations, and the Internet, the cost of a connection can be measured by other attributes such as the effort expended in maintaining a relationship, or the time spent in accessing a web page. In all of these networks, the associated cost metric is not necessarily a constant over all connections. Although various cost metrics influence networks, in this paper we investigate only a physical metric and ask how placing a *cost* on the length of an edge affects the connectivity of the network. However, we work with the motivation that the results are also applicable to networks with different underlying cost metrics.

Although recent work showed small worlds to be pervasive in a range of networks that arise from both natural and manmade technology, the hows and whys of this ubiquity have not been explained. The fact that small worlds seem to be one of nature's ''architectural'' principles leads us to ask what constraints might force networks to choose a small-world topology. We investigate whether the emergence of a small-world topology, in networks where the physical distance is a criterion that cannot be ignored, can arise as a tradeoff between maximal connectivity and minimal wiring.

We now briefly describe the small-world model of Watts and Strogatz (WS), and also introduce the notation that we shall use. Watts and Strogatz considered a ring lattice, which is n vertices arranged at regular intervals on a ring, with each vertex connected to its k nearest neighbors. Disorder is introduced into the graph by randomly rewiring each of the edges with a probability p. While at $p=0$ the graph remains k regular, at $p=1$ a random graph results [3]. At intermediate p, small-world graphs result. Watts and Strogatz quantified the structural properties of graphs by two parameters, the *characteristic path length* L and the *clustering coefficient* C. The characteristic path length reflects the average connectivity or vertex-to-vertex separation of the network, while the clustering coefficient measures the extent to which neighbors of a vertex are neighbors of each other. They found that their marginally randomized graphs are symptomatic of the world in which we live, in that they are characterized not only by a small degree of separation between vertices, but are also highly clustered. Small-world graphs, therefore, by locally looking like regular graphs but globally behaving like random graphs, reconcile the properties of the graph topologies at the two extremes. Finally we also point out that two kinds of distance measures can be employed in a graph. One is the *graph* distance, the minimal number of links between any two vertices of the graph; and the other is the *physical* distance between these vertices. We use Euclidean distance as a

*Email address: nisha@csa.iisc.ernet.in
†Present address: G. R. Harrison Spectroscopy Laboratory, Rm. 6-014, MIT, Cambridge, MA 02139.

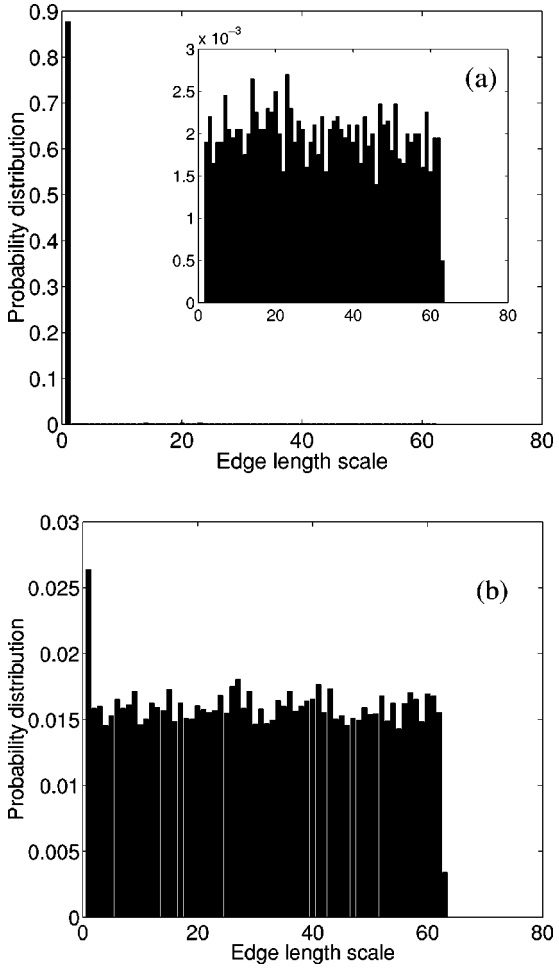

FIG. 1. Edge scale distribution at (a) $p=0.125$, and (b) $p=1.00$, p being the amount of disorder introduced into the $n=250$, $k=4$ regular network using the WS rewiring procedure to introduce small-world behavior. The inset in (a) displays the distribution of all length scales other than the unit scale. Both plots are averaged over 25 samples.

metric of the latter since, it is a natural choice for physical networks, noting, however, that other suitable metrics can be constructed for nonphysical networks.

Section II elaborates on our motivations for questioning whether small worlds can arise as a result of an optimization between conflicting constraints. In Sec. III we then describe our optimization model. Section IV follows with a study of the resulting optimized network topology, while Sec. V highlights differences between the WS model and the optimization model with respect to their small-world characteristics. In Sec. VI we ask whether a similar topology can be found in existing natural or manmade networks. Finally, we conclude in Sec. VII.

II. CAN SMALL WORLDS ARISE AS THE RESULT OF AN OPTIMIZATION?

Consider a toy model of the brain. Let us assume that it consists of local processing units, connected by wires. What constraints act on this system? On the one hand, one would want the highest connectivity (shortest path length) between the local processing units, so that information can be exchanged as fast as possible. On the other hand, it is wasteful to wire everything to everything else. The energy requirements are higher, more heat is generated, more material needs to be used, and, consequently, more space is occupied. Unrealistic though this model is, it motivated us to examine whether small worlds would emerge as the result of these constraints.

Cherniak investigated the influence of wiring as a constraint on neural structure. In Refs. [5,6] he questioned whether neural networks optimize the positioning of their components due to a constraint on the wiring of connections between them. He found that the structure of a nervous system at various hierarchical levels supports the hypothesis that component placement optimization is a driving principle of neuroanatomy.

The concept of multiple scales was introduced by Kasturirangan [7], where he asserted that the fundamental mechanism behind the small-world phenomena is not disorder or randomness, but the presence of edges at many different length scales. He defined the *length scale* (or *range*) of a newly introduced edge e_{ij}, to be the graph distance between its adjacent vertices i and j *before* the edge was introduced. He further emphasized that the distribution of length scales of the new edges is of significantly more consequence than whether the new edges are long, medium, or short ranged.

Following Kasturirangan [7], we introduce the *edge scale distribution* of a graph to be simply the histogram of the length scales of all its edges. To obtain the distribution of edges in a graph whose edges are merely rewired (without the introduction of additional edges), we consider the length scale of a rewired edge to be the distance between its adjacent vertices in the corresponding regular graph. Starting with a k-regular graph and using the WS rewiring procedure, we study the edge scale distribution at several values of the rewiring parameter p. Figure 1 shows the edge scale distribution at two values of p. The edge scale distribution in the small-world regime ($p=0.125$), is shown in Fig. 1(a). Due to the introduction of a small amount of disorder, a few edges are rewired to become far, and consequently have a large length scale. However, they are too few in number to significantly alter the edge scale distribution, and hence the edges of unit length scale dominate the distribution. Figure 1(b) shows the edge scale distribution at $p=1$, a random graph. Here the edges are uniformly distributed over the entire length scale range, viz. from 1 to n/k. The network still retains a slight bias toward the unit length scale. At both these values of randomness, however, the characteristic path length scales logarithmically with n. There thus appears to be some factor that constrains the distribution of edge length scales to (a) and not (b), namely, restricting the rewiring to just a few far edges. We question whether the association of a cost to each edge proportional to its length, serves to work as this constraint.

III. OPTIMIZATION MODEL

We use the Metropolis algorithm [8,9] for simulated annealing to find a network which results in the best optimiza-

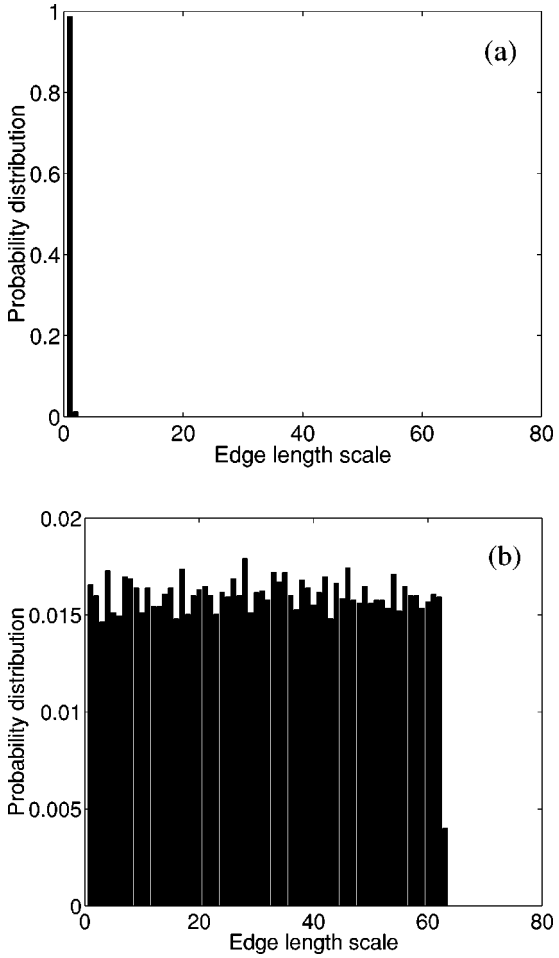

FIG. 2. Edge scale distribution resulting from optimization at (a) $\lambda=0$, and (b) $\lambda=1$ for a network having $n=250$ and $k=4$. Both distribution plots are averaged over 25 simulations.

tion of the objective function E, whose minimization is the goal of the procedure. The network used in the model is that of vertices arranged symmetrically along a ring. The size of the network n, as well as the total number of edges, is fixed. So also are the positions of the vertices, which are equally spaced along the circumference of the circle. Initially, the network is k regular, similar to the WS model. The network configuration has an associated energy E, a function of both its wiring cost and the average degree of separation between its vertices. The objective function E is taken to be

$$E=\lambda L+(1-\lambda)W, \tag{1}$$

a linear combination of the normalized characteristic path length L and the normalized wiring cost W. The characteristic path length L, as defined by Watts and Strogatz, is the average distance between all pairs of vertices, given by

$$L=\frac{1}{n(n-1)}\sum_{i\neq j} d_{ij}, \tag{2}$$

where d_{ij} is the number of links along the shortest path between vertices i and j. It is therefore a measure based on *graph* distance, and reflects the global connectivity among all vertices in the graph. The wiring cost W, in contrast, is a measure of the *physical* distance between connected vertices. The cost of wiring an edge e_{ij}, is taken to be the Euclidean distance between the vertices i and j. Hence the total wiring cost is

$$W=\sum_{e_{ij}} \sqrt{(x_i-x_j)^2+(y_i-y_j)^2}, \tag{3}$$

where (x_i,y_i) are the coordinates of vertex i on the ring lattice. The characteristic path length L is normalized by $L(0)$, the path length in the k-regular network; W is normalized by the total wiring cost that results when the edges at each vertex are the longest possible, namely, when each vertex is connected to its diametrically opposite vertex, and to the vertices surrounding it. The parameter λ is varied depending on the relative importance of the minimization of L and W. One can regard $(1-\lambda)$ as the wiring cost per unit length, and W as the length of wiring required. This model therefore includes a constraint on both graph distance and physical distance, resulting in graphs that combine both relational as well as spatial mechanisms [10].

Our model optimizes the interconnectivity among components in networks when component placement is fixed. By contrast, Cherniak investigated whether the positioning of components is optimized in networks, when their component interconnectivity is fixed. Both works, therefore, though built on the hypothesis that wiring needs to be minimum, follow contrasting approaches to optimality.

Starting from the initial regular network, a standard Monte Carlo scheme [8] was used to search for the energy minimum. Similar to the WS model, duplicate edges and loops were not allowed, and it was ensured that the rewiring did not result in isolated vertices. The starting value for T, the annealing temperature, was initially chosen to be the initial energy E. The temperature was then lowered in steps, each amounting to a 10% decrease in T. Each value of T was held constant for $150n$ reconfigurations, or for $15n$ successful reconfigurations, whichever was earlier.

IV. OPTIMIZED NETWORKS: RESULTS

Since minimum characteristic path length, and minimum wiring cost are contradictory goals, the optimization of either one or the other will result in networks at the two ends of the randomization spectrum. As expected, at $\lambda=0$, when the optimization function concentrates only on minimizing the cost of wiring edges, a regular network emerges with a uniform degree and a high characteristic path length ($L\sim n$). The edge scale distribution shows all edges to be concentrated almost entirely within the unit length scale, as shown in Fig. 2(a). At $\lambda=1$, when only the characteristic path length is to be minimized, the optimization results in a ''random'' network ($L\sim\log n$). The edge scale distribution shown in Fig. 2(b) has edges having lengths distributed uniformly over the entire length scale range. A later section explains how, despite the logarithmic dependence of L on n, and the flat edge scale distribution, the network at $\lambda=1$ is not the same as that of a random graph in Ref. [4].

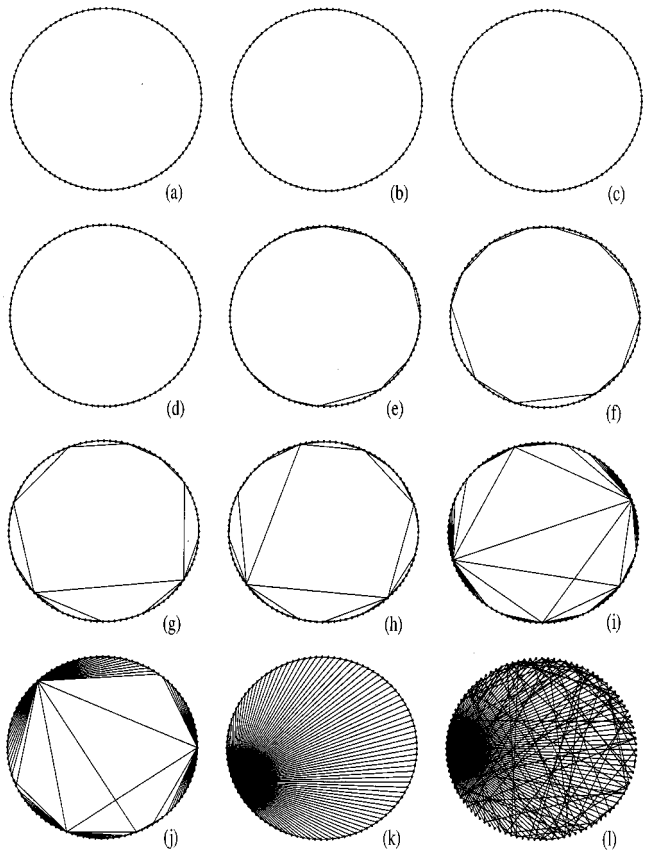

FIG. 3. The ring lattice displays illustrate the evolution of hubs as λ is varied over the [0,1] range for an $n=100$, $k=4$ optimized network. Very short interhub links cannot be distinguished apart from local vertex connectivity; however, longer range interhub links are clearly visible. Distinct hub centers illustrate the presence of hubs, as well as their variation in size and number. The single hub center at the universal hub limit is clearly illustrated. λ: (a) 0.0, (b) 5×10^{-4}, (c) 5×10^{-3}, (d) 1.25×10^{-2}, (e) 2.5×10^{-2}, (f) 5×10^{-2}, (g) 1.25×10^{-1}, (h) 2.5×10^{-1}, (i) 5×10^{-1}, (j) 7.5×10^{-1}, (k) 7.8×10^{-1}, and (l) 1.0.

A. Emergence of hubs

At intermediate values of λ the optimization model results in *hubs*, that is, a group of nodes connected to a single node [7,11,12]. The emergence of hubs is due to the contribution of L to the optimization function. Due to the constraint which seeks to minimize the physical distance between connected vertices as well, hubs are formed by vertices close to one another. In addition, the minimization of the graph distance ensures the existence of connections between hub centers, enabling whole hubs to communicate with each other. The edges at any hub center therefore span a wide range of length scales.

The extreme situation is a *universal hub* [7]: a single node, with all other nodes having connections to it. However, except for situations when the cost of wiring is negligible, we find that the optimization does not result in a universal hub. This is apparent since a universal hub requires all the remaining $n-1$ vertices to have connections to the vertex at the center of the hub. This results in length scales which span the entire scale range, with long connections being prohibitively expensive. A real world example of such a universal hub network is unlikely, since a large hub is a bottleneck to traffic through it, resulting in overcrowding at the hubs [7]. Hence the need for multiple, and consequently smaller, hubs.

Watts [10] defined the *significance* of a vertex v as the characteristic path length of its neighborhood $\Gamma(v)$, in the absence of v. Hub centers are significant, since they contract distances *between* every pair of vertices within the hub. Thus vertex pairs, although not directly connected, are connected via a single common vertex. Hence the average significance, a measure which reflects the number of contractions, is considerable. Thus, in contrast to the WS model, where networks become small due to shortcuts, here smallness can be attributed to the small fraction of highly significant vertices.

The formation of the universal hub at sufficiently large λ is not surprising, since it can be shown that for a network that minimizes L and employs only rewirings, a universal hub will effect the largest minimization. The formation of *multiple* hubs, however, is due to the role played by W in the optimization, which is to constrain the physical length of edges, and therefore the size of hubs. As the hubs grow, whenever the cost of edges from the hub center to farthest nodes become high, the edges break away, resulting in multiple hubs. Thus the high wiring cost prevents the formation of very large hubs, and controls both the size and number of hubs.

Figures 3 and 4 demonstrate the evolution of hubs in an $n=100$, $k=4$ optimized network as λ is varied between 0 and 1. While Fig. 3 uses ring-lattice displays to illustrate the evolution, Fig. 4 illustrates the same networks as two-dimensional (2D) displays. In the ring-lattice displays, vertices are fixed symmetrically around the lattice, with hub centers and long-range interhub links being clearly visible. The 2D displays are generated by a graph drawer which uses a spring embedder to clearly demonstrate vertex interconnectivity. As vertices are no longer fixed along a ring lattice, short-range interhub links can be distinguished apart from local connectivity. We point out that the figures show typical network topologies at different λ's. Since the annealing was done sufficiently slowly, all hub networks obtained at each value of λ using different random number seeds are similar in topology, and more quantitatively, their small-world properties and edge scale distributions have no significant variations.

B. Hub evolution

We now detail the evolution of hubs using the edge scale distribution shown in Fig. 9, and hub variations described in Figs. 3 and 4. All three figures show the same $n=100$, $k=4$ network at various λ.

In Figs. 9(a) and 4(a), optimization results in a near regular network, with hardly any hubs. When the cost reduces very slightly to allow for a slight increase in edge wiring, very small hubs are formed. Hence for increasing but very small λ, Figs. 9(b) and 9(c) show the edges to be almost entirely concentrated in the unit length scale, with very few longer edges. The nonunit length scale edges account for

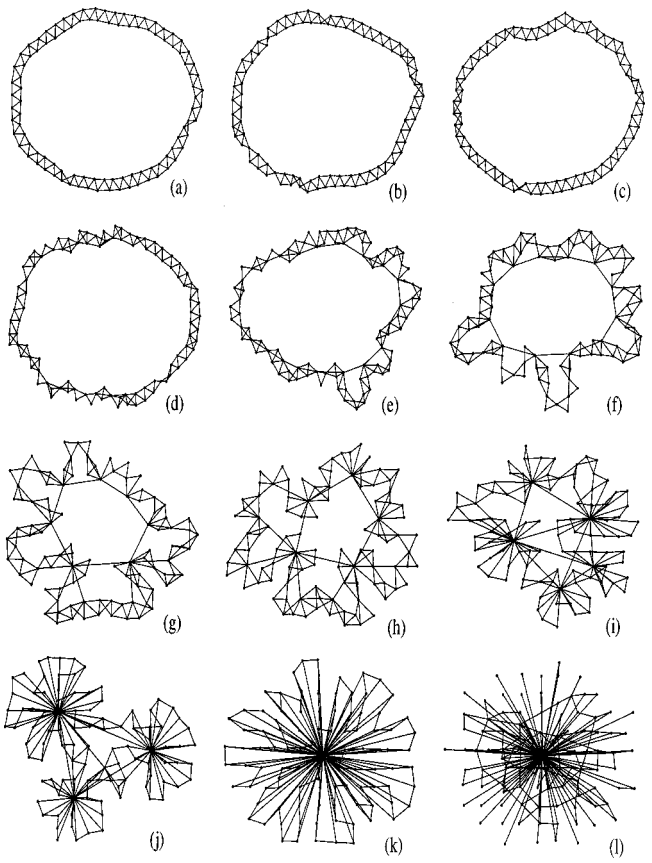

FIG. 4. This figure illustrates the evolution of hubs for an $n=100$, $k=4$ optimized network as λ is varied over the same [0,1] range as the previous figure. The same networks are displayed as 2D graphs using a graph generator with a spring embedder. Now, since vertices are no longer displayed as being fixed along a ring lattice, vertex interconnectivity, as well as the emergence of hubs and their variation in size and number is well illustrated. λ: (a) 0.0, (b) 5×10^{-4}, (c) 5×10^{-3}, (d) 1.25×10^{-2}, (e) 2.5×10^{-2}, (f) 5×10^{-2}, (g) 1.25×10^{-1}, (h) 2.5×10^{-1}, (i) 5×10^{-1}, (j) 7.5×10^{-1}, (k) 7.8×10^{-1}, and (l) 1.0.

very few and very small hubs, as illustrated in Figs. 4(b) and 4(c). Due to their small size and very short interhub links, the hubs are indistinguishable from local vertex connectivity in the ring lattice displays in Figs. 3(a)–3(d).

A slight fall in the wiring cost permits an increased number of hubs. However, the cost of wiring continues to be high enough to constrain hubs to still be rather small. The distribution of scales in Fig. 9(d) hence still shows only two length scales, but with a marked increase in edges of the second length scale. The effort toward minimizing L ensures that the few hubs are bunched close together, so that short interhub links can be used to enable the maximum distance contraction possible [Fig. 4(d)].

When further reduction in cost permits increased wiring, it is mostly the interhub links that take advantage of the reduced cost to enable hubs to be scattered over the entire network. Figure 9(e) shows clearly the multiple length scales generated by interhub links. The marked increase in the range of the interhub links [Fig. 3(e)] allows them, for the first time, to be visible in the ring-lattice plots. Figure 4(e) shows that there is not much variation in the hub size, except for the longer range of the interhub links.

Figures 9(f)–9(h) and 3(f)–3(h) demonstrate that, as λ increases further, the length and number of far edges are progressively less constrained, the extended length permitting larger and many more hubs. Vertices lose their local nearest-neighbor interconnectivity as hub centers dominate in connectivity [Figs. 4(f)–4(h)]. However, as the size of hubs progressively increases, they are consequently reduced in number [Figs. 4(i)–4(k)]. The density of interhub links increases though to yield greater interhub distance contraction. In Figs. 4(i)–4(k), one also observes efforts toward a uniform reduced local connectivity.

Figures 4(i) and 4(j) are marked by a sharp reduction in the number of hubs as the hubs balloon in size. This evolution culminates in the emergence of the universal hub [Fig. 4(k)], a single hub of connectivity. The formation of edges between the hub center and all the other $n-1$ vertices, as illustrated in Fig. 3(k), results in a uniform distribution of nonunit length scale edges. Wiring, which is still associated with a cost, albeit small, ensures that the remainder of the edges are entirely local, as can be observed from the distribution in Fig. 9(k). Even when the cost of wiring drops further, the universal hub topology continues to be retained, since its value is sufficient to constrain the remainder of the edges to be still entirely local.

Only when the cost of wiring is zero or is negligibly small, does a change in topology result. Figures 3(l) and 4(l), at $\lambda=1$, demonstrate that although the universal hub is retained [13], due to the absence of any effort towards minimal wiring, edges are uniformly distributed across the entire length scale range, as shown in Fig. 9(l). The loss in local connectivity can be clearly seen in comparison to Figs. 3(k) and 4(k). Optimization toward minimizing only L results in the *re-emergence* of multiple hubs, but the removal of the constraint on wiring allows hubs to be composed of largely nonadjacent vertices.

In summary, during the evolution of hubs illustrated in Figs. 3, 4, and 9, as the cost of wiring is decreased, the following sequence is seen: (i) hubs emerge, and grow in size and number; (ii) they increase in the range and density of interhub links; (iii) there is a subsequent reduction in the number of hubs; (iv) there is the formation of a universal hub; and (v) hubs re-emerge, accompanied by a loss in local vertex interconnectivity, while the universal hub remains.

V. OPTIMIZATION AND THE WS MODEL: SOME COMPARISONS

In this section we present further results but against the backdrop of the WS model. Recollecting, to define small-world behavior, two ingredients were used by Watts and Strogatz. The first was the characteristic path length L, a global property of the graph, while the second was the clustering coefficient C, a local property which quantifies neighborhood ''cliquishness.'' Associated with each vertex v is its neighborhood Γ_v, the k_v vertices to which it is directly connected and among which there can be a maximum of $k_v(k_v$

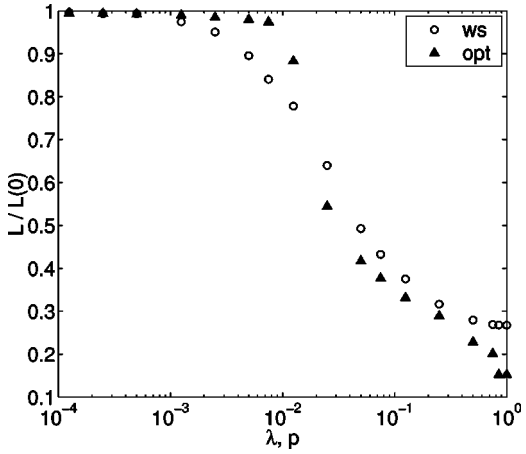

FIG. 5. Variation in the normalized characteristic path length $L/L(0)$, vs p and λ, for the WS model and optimization model, respectively. Both characteristic plots are obtained for a $n=100$, $k=4$ network, and averaged over 40 simulations.

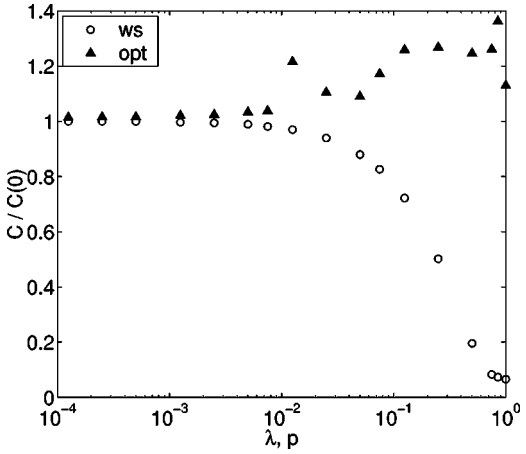

FIG. 6. Comparison between the WS model and the optimization model for a $n=100$, $k=4$ network, with respect to the variation in their normalized clustering coefficient $C/C(0)$. Both variations are averaged over 40 realizations.

$-1)/2$ connections. C_v, the clustering coefficient of v, denotes the fraction of the links actually present among its neighbors. It is defined as

$$C_v=\frac{|E(\Gamma_v)|}{\binom{k_v}{2}}, \tag{4}$$

and C is C_v averaged over all v.

The WS and optimization models are compared with respect to their normalized small-world characteristics. In addition, we study their different behaviors with respect to normalized wiring and degree, as well as edge scale distribution. All results are obtained using an $n=100$, $k=4$ network. Each plot is the result of averaging over 40 simulation runs.

A. Characteristic path length

We begin our comparison with the characteristic path length, the parameter whose smallness gives these networks their name. Figure 5 compares L for the WS and optimized models. The control parameters in the two models, the optimization parameter λ and the WS parameter p, are similar in that they both control the introduction of far edges. It should be remembered, however, that while p controls only the *number* of far edges, allowing their length scales to be uniformly distributed across the entire range, λ constrains not only the number but also the physical *length* of far edges.

In both cases, L shows a sharp drop that signifies the onset of small-world behavior. However, in contrast to the gradual drop effected by the random assortment of rewired edges in the WS model, the drop due to hub formation is much sharper. Although its initial reduction is smaller due to the additional constraint on edge length, its final value is much lower than the WS random graph limit.

The variation in L resulting from optimization can be understood from the role played by the hub centers in contracting distance between pairs of vertices. Before the cliff, the hubs, being few and very small, effect a very slight distance contraction. The tip of the cliff forms due to a marked increase in small hubs, while a sharp drop occurs when extended range interhub links yield a pronounced distance contraction between many distant hubs and their widely separated neighborhoods. The transition from many small hubs to much larger and consequently fewer hubs results in a gradual reduction in L. Finally, at the emergence of the universal hub, which has no counterpart in the WS model, the single hub center contracts the distance between *every* pair of vertices, resulting in an average distance less than 2.

B. Clustering coefficient

Figure 6, which compares the variation in the clustering coefficient for the two models, shows a far more interesting behavior. The drop in local connectivity that is seen in the WS model does not occur *at all* for the optimized network because of the formation of hubs. Although the clustering coefficient was not a characteristic that we sought to maximize, high cliquishness emerges. Figure 6 shows that the formation of hubs sustains the clustering coefficient at a value higher than that for the corresponding regular graph, unlike the WS model. Thus the similarity between p and λ as control parameters is only valid for L.

Before we make a more detailed analysis of Fig. 6, we discuss the clustering coefficient further. For a vertex v, its *neighborhood size* k_v, plays a significant role toward the value of its clustering coefficient C_v. The smaller k_v is, the smaller the number of possible intraneighborhood edges. Hence vertices which lose in connectivity gain in cliquishness. In a similar manner, vertices which gain in connectivity lose in cliquishness because of their larger neighborhood size. This is because, although the vertices have a larger number of intraneighborhood edges, they form a smaller fraction of the total number of possible edges. At the universal hub limit, the hub center has the least clustered neighborhood owing to the fact that all the remaining $n-1$ vertices form its neighborhood. The clustering coefficient can be determined to be approximately $(k-2)/n$. Although the aver-

age degree remains unchanged, the varying hub size and number can influence which neighborhoods dominate the average clustering coefficient.

In addition, a factor which influences the *clustering* within the neighborhood of a vertex is the inclusion of a hub center to the neighborhood. The effect on clustering differs depending on the range of the link between the vertex and the hub center. If the range of the link is large and the vertex lies outside the hub, then a far away hub center is included into the vertex's otherwise locally connected neighborhood. The hub center has little or no association with the remaining neighbors of the vertex, and so it lowers the average cliquishness. However, if the vertex lies within the hub, the hub center serves as a node which is connected to all or a large fraction of the neighbors. Hence the neighborhood of the vertex becomes more clustered. This effect is more pronounced when (1) the size of the neighborhood is small, (2) the vertex includes more than one hub center in its neighborhood, and (3) the hub whose center is being included is composed of largely local neighbors. Thus, unlike L, which is tuned by a single parameter, C is controlled by many more; a point that we will return to shortly in the analysis of Fig. 6.

From Fig. 6 we see as expected for the WS model, that $C(p)$ falls as approximately $C(0)(1-p)^3$ [14], and eventually drops almost to zero for the completely randomized graph. This is due to the increasing number of inclusions of random far nodes into otherwise locally connected neighborhoods. In contrast, the presence of hubs in the optimization model ensures that $C(\lambda)$ never falls below $C(0)$, reaching its maximum when the network converges to a universal hub.

Keeping in mind the evolution of the optimized network shown in Figs. 3 and 4, we can qualitatively understand the behavior of $C(\lambda)$ in Fig. 6. When λ is small, the network is dominated by regular neighborhoods [Figs. 3(a)–3(c) and 4(a)–4(c)]. Since hub centers gain in connectivity at the expense of adjacent vertices, vertices adjacent to hub centers gain in cliquishness due to their reduced neighborhood sizes, as described earlier. Despite there being just a few small hubs, since there are more reduced connectivity vertices than hub centers, the average C is raised slightly above that of a regular graph.

With a slight increase in λ, Figs. 3(d) and 4(d) show a sharp increase in C. At this point, the marked increase in hubs, with only a slight increase in size, results in a pronounced increase in the number of reduced connectivity vertices. Many of these vertices have neighborhoods which are completely clustered (called cliques), since in addition to their reduced size, they include one or more hub centers into their neighborhood. Cliques, not surprisingly, dominate the average, resulting in a large jump in C. However, the emergence of long range interhub links in Figs. 3(e) and 3(f) and 4(e) and 4(f) results in lowering C. Their introduction causes: (1) hub centers to have lowered cliquishness owing to the inclusion of distant nodes into their neighborhoods; and (2) some reduced connectivity neighborhoods to no longer be complete cliques due to the inclusion of the center of a hub, which has lost local neighbors to interhub neighbors.

In Figs. 3(g) and 3(h) and 4(g) and 4(h), C again rises due to increased cliques generated not only by the increased number and size of hubs, but also because their larger size allows for the inclusion of local neighbors once again. However in Figs. 3(i)–3(k) and 4(i)–4(k), while the few hubs gain in connectivity (and consequently lose in cliquishness), the remaining vertices veer toward a uniform reduced connectivity. The resulting marked reduction in the number of cliques accounts for the slight drop at Figs. 3(i) and 4(i), while the subsequent near uniform reduced connectivity serves to further raise C. Finally, at the universal hub limit [Figs. 3(k) and 4(k)], all vertices have a uniform reduced vertex connectivity at the expense of the single hub center. Having gained in cumulative connectivity, the hub center has a very low clustering coefficient of approximately $(k-2)/n$. However, the remaining reduced sized neighborhoods, and their inclusion of the hub center, ensure the average C shoots up to its maximum.

In Figs. 3(l) and 4(l), the average clustering falls due to the nonuniformity in vertex connectivity. However the inclusion of the universal hub center into every neighborhood ensures that C does not drop too much. The multiple hubs result in a variation in vertex connectivity, with hub centers gaining in connectivity at the expense of others. This leaves a few vertices having a *single* connection. With a neighborhood of only 1, and no intra-neighborhood connectivity, these vertices are totally unclustered [15], which accounts for the drop in C.

Thus we see an interesting interplay between neighborhood size, hub center inclusions, and the number and range of interhub links. However, the data for the variation of C in the optimized model are noisy, mainly because k is very small. Constraints in computational resources have forced us to work with small n. Further, to maintain the sparseness condition of $n \gg k$, a low k was used, which does not really satisfy the WS condition that $k \gg 1$. Due to the small k, even a small loss in connectivity, can cause neighborhood cliquishness to rise sharply. Although different factors come into play during the C variation, the spikes are due to the pronounced effect of reduced connectivity neighborhoods, and in particular to those of cliques. The cliques serve to maintain the entire C variation higher than would probably result for higher k. Work is in progress to obtain data using large k networks.

C. Wiring and degree

Figure 7(a) displays the increase in the cost of wiring, or alternatively, the amount of wiring, with p and λ. The comparison between the optimization model and the WS model illustrates clearly the difference made by the inclusion of the minimal wiring constraint. For small λ, both models exhibit similar wiring cost. At larger λ however, the absence of a similar constraint in the WS model results in a much greater amount of wiring. The clear advantage exhibited by the optimized networks persists until $\lambda = 1$, when optimization neglects the minimization of wiring cost entirely. At this point, the optimized network uses greater wiring than its WS counterpart, but only slightly.

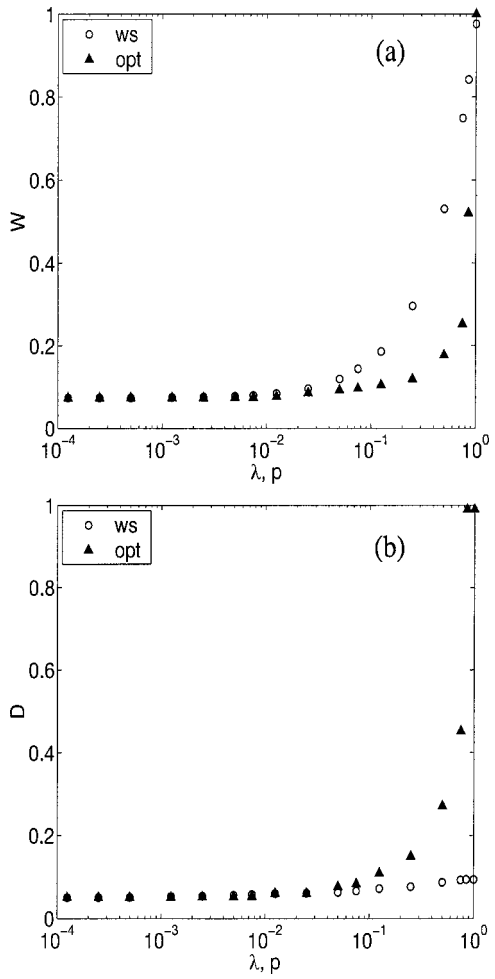

FIG. 7. Comparison between the WS and optimization models vs p and λ, respectively. (a) Variation in the wiring cost W, normalized by the optimized value of $W(1)$. (b) Variation in the maximum degree D, normalized by network size n. D is equivalent to the size of the largest hub in the optimization model. Each plot shows results for a $n=100$, $k=4$ network averaged over 40 samples.

In contrast to the WS model, a constraint on degree is not maintained in the optimized model. Figure 7(b) shows how the maximum degree increases with λ and p, for the two models. The maximum degree D, is normalized by the network size n. For the optimized network, D is equivalent to the size of the largest hub. At small λ, there is no difference between the two models, but once hubs begin to emerge, D increases sharply for the optimized network. At the universal hub and the near random graph limit, the maximum degree is $n-1$, the size of the universal hub. In contrast, each edge, being rewired only once in the WS model, allows for only a slight variation in degree. One also observes a similarity in the variations of W and D for the optimized networks, since W controls the size of hubs, and hence D.

D. Edge scale distribution

A comparison between the WS (Fig. 8) and optimized (Fig. 9) edge scale distributions clearly illustrates the differing underlying mechanisms that result in small worlds. Since the WS rewiring mechanism exercises no restraint on the length scales of the rewired edges, the rewired edges are correspondingly uniformly distributed over the entire length scale range. In contrast, for the minimally wired networks, lower length scales occur with a higher probability.

Figure 10 shows plots of the optimized edge scale probability distribution on a log-log scale, where a power-law behavior is seen. Figures 10(a) and 10(b) illustrate the edge scale distributions at varying λ, while Fig. 10(c) is a combined plot which demonstrates the variation in the power-law distributions with λ. Each distribution is displayed along with its associated linear least-squares fit. The variation in their exponents, as obtained from the linear least-squares fit to the data against λ, is shown in Fig. 10(d).

The variation in the power-law exponents with λ can be clearly demarcated into two regions. The first, spanning two orders of magnitude variation in λ, exhibits a very slight exponent variation. Figure 10(a) illustrates the typical probability distribution in this regime. Just two points emerge, since the high wiring cost constrains almost all edges to have a unit length scale, with a very slight probability of a higher length scale. A sharp jump in the exponent marks the beginning of the second regime. Figure 10(b) illustrates the typical probability distribution in this regime. It is seen that a straight line is a reasonably good fit to the data over a wide edge scale range. Finally, when $\lambda=1$, and a near random network is achieved, a flat distribution of length scales results with each scale being equally probable. The combined plot of all the distributions with their associated least-squares fits, although noisy, illustrates the behavior of the data [Fig. 10(c)].

The exponent variation clearly reveals two regimes of behavior. The first jump in the exponent corresponds to the onset of small-world behavior, the first perceptible reduction in L that is seen in Fig. 5. This marks the beginning of the multiple scale regime, and is also a signature of hub formation. As mentioned previously, due to computational constraints we were unable to investigate larger networks. We believe that the noise in the data is due to the small size of the networks that we have studied.

Finally, we wish to comment upon Kasturirangan's multiple scale hypothesis. Figure 10(d) clearly demonstrates the connection between the onset of small-world behavior and the emergence of multiple length scales in the network. This clearly supports the claim in Ref. [7] that small worlds arise as a result of the network having connections that span many length scales. These are the first numerical results in support of his hypothesis. The power-law behavior also demonstrates that any optimized distribution of multiple scales will result in a hierarchy of scales. It is to be noted that multiple scales contribute to reducing L in the WS model as well. However, since no restriction on the length of edges exists, *any* nonzero p will result in multiple scales. Hence, the onset of small-world behavior appears with a smooth reduction in L.

We also observe a power-law tail for vertex degree. We find that most vertices have a small degree, and some are

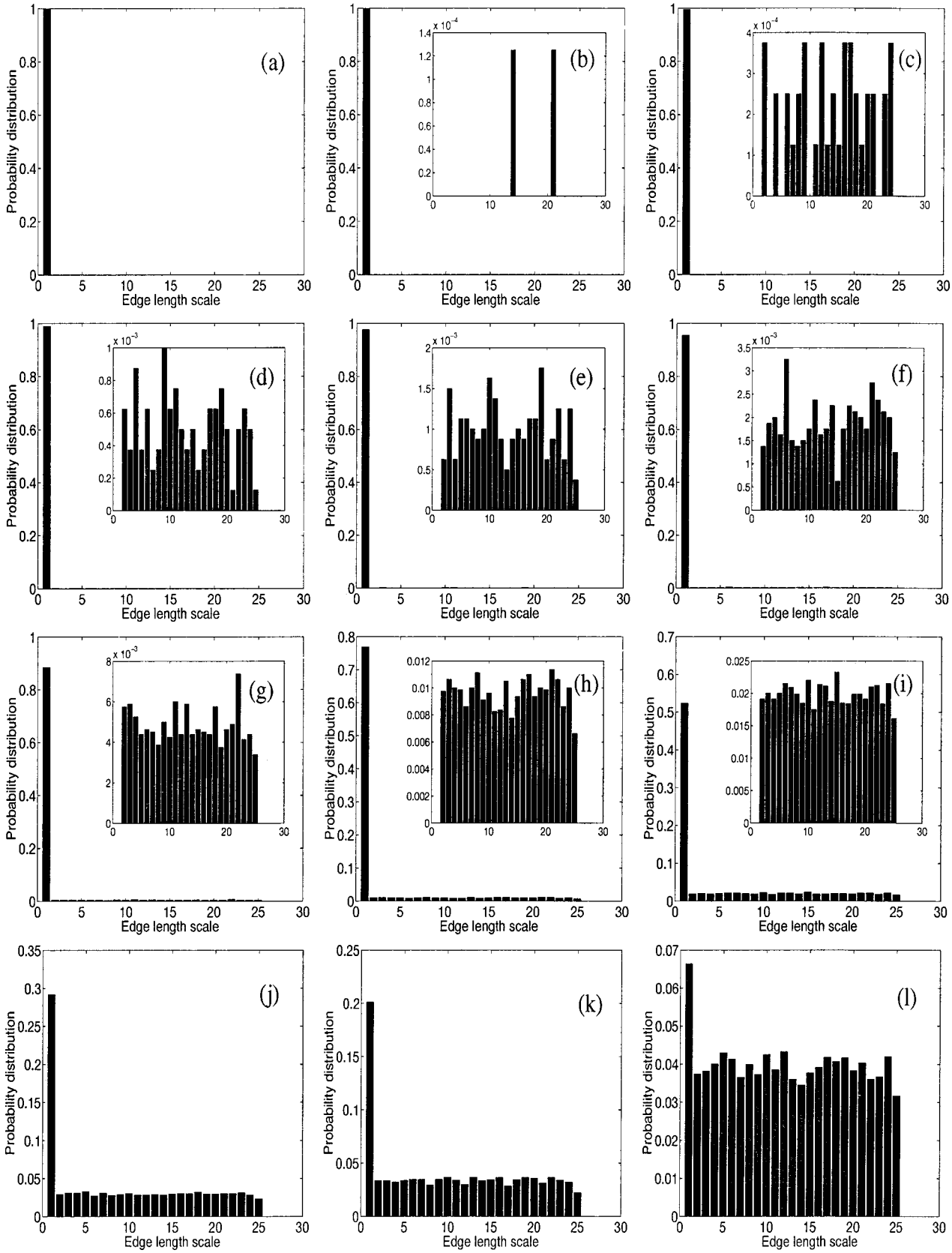

FIG. 8. Edge scale distribution for an $n=100$, $k=4$ network randomized using the WS rewiring procedure at various values of p: (a) 0.0, (b) 5×10^{-4}, (c) 5×10^{-3}, (d) 1.25×10^{-2}, (e) 2.5×10^{-2}, (f) 5×10^{-2}, (g) 1.25×10^{-1}, (h) 2.5×10^{-1}, (i) 5×10^{-1}, (j) 7.5×10^{-1}, (k) 7.8×10^{-1}, and (l) 1.0. The inset in each plot shows the distribution of all scales with the unit length scale excluded. Each plot is an average over 40 simulations.

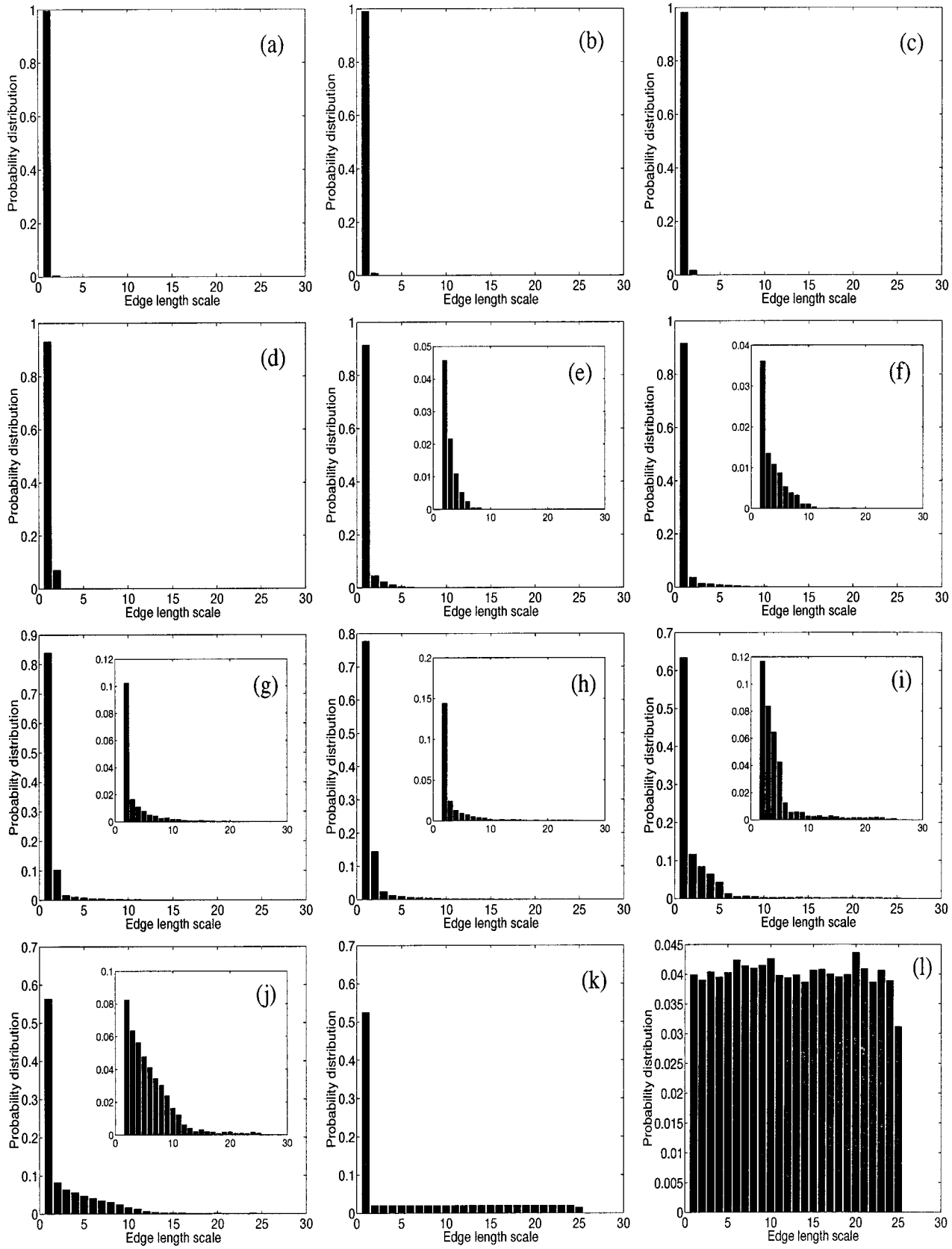

FIG. 9. Edge scale distribution for an $n=100$, $k=4$ network optimized at various values of λ: (a) 0.0, (b) 5×10^{-4}, (c) 5×10^{-3}, (d) 1.25×10^{-2}, (e) 2.5×10^{-2}, (f) 5×10^{-2}, (g) 1.25×10^{-1}, (h) 2.5×10^{-1}, (i) 5×10^{-1}, (j) 7.5×10^{-1}, (k) 7.8×10^{-1}, and (l) 1.0. The inset in each plot shows the distribution of all scales with the unit length scale excluded. Each plot is an average over 40 simulations.

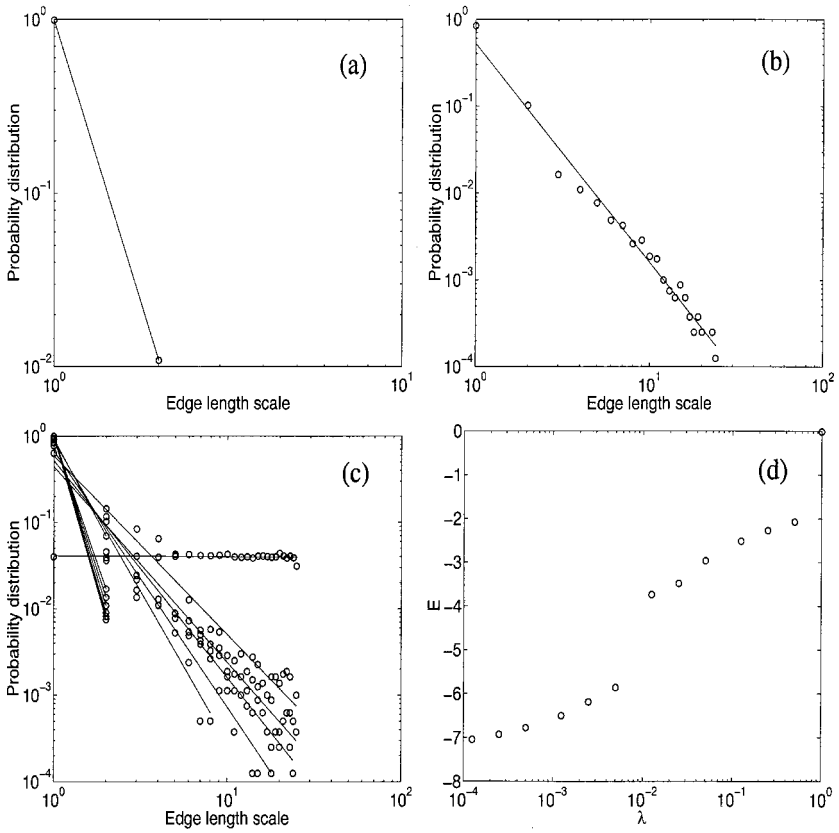

FIG. 10. Log-log plot of the edge scale probability distribution at (a) $\lambda = 1.25 \times 10^{-3}$, and (b) $\lambda = 1.25 \times 10^{-1}$, for an $n = 100$, $k = 4$ optimized network. (c) Combined plot of the probability distributions, each with its associated linear least-squares fit. (d) Variation in the power-law exponents with λ. Each distribution plot is averaged over 40 realizations.

well short of the average degree, with vertices at hub centers gaining at their expense. Owing to the small network size, however, the scaling range is rather limited, and so we do not include these results.

VI. DO SIMILAR NETWORKS EXIST?

Any efficient transportation network works under a similar underlying principle of maximizing connectivity while ensuring that the cost is minimized. Our results seem to indicate that any efficient transportation network will be a small world, and in addition will exhibit a similar hub connectivity. In a clear illustration of the underlying principle, any map of airline routes or roadways shows big cities as being hubs of connectivity. This is hardly surprising, however, since in such networks a conscious effort is made toward such a minimization. However, the same philosophy may well be at work in natural transportation and other biological networks.

Although this model concentrates on natural and artificial transportation networks, which are networks composed of links that possess a measurable physical length allowing us to associate a cost with each edge, we would also like to point out that our observed hub structure can be seen in a number of large complex networks ranging from fields as diverse as the World Wide Web to the world of actors. Kleinberg *et al.* [16] observed the following recurrent phenomenon on the Web: for any particular topic there are hubs of information as well as connectivity, constructed by a dual interlinking. ''Authoritative'' pages focused on the topic are recognized by the many links pointing to them, while ''hub'' pages are identified by the many links pointing from them to pages rich in useful material pertinent to the topic. In addition, Barabási and Albert [17] explored several large databases describing the topology of large complex networks that traverse a host of disciplines. They observed that, regardless of the system and the nature of its components, the probability $P(k)$ that a vertex in the network interacts with k other vertices decays as a power law following $P(k) \sim k^{-\gamma}$. The power law for the network vertex connectivity indicates that highly connected vertices (large k) have a large chance of occurring, dominating the connectivity, and hence demonstrating the presence of hubs in these networks. Further, it was noted in Ref. [18] that the small-world phenomenon in the world of actors arises due to ''linchpins,'' hubs of connectivity in the acting industry that cut across genres and eras. Thus hubs seem to constitute an integral structural component of a number of large and complex random networks, both natural and manmade.

VII. CONCLUSIONS

Watts and Strogatz showed that small worlds capture the best of both graph worlds: the regular and the random. However, there has been no work citing reasons for their ubiquitous emergence. Our work is a step in this direction, questioning whether small worlds can arise as a tradeoff between optimizing the average degree of separation between nodes in a network, as well as the total cost of wiring.

Previous work concentrated on small-world behavior that arises as a result of the random rewiring of a few edges with no constraint of the length of the edges. On introducing this constraint, we have shown that an alternate route to small-world behavior is through the formation of hubs. The vertex at each hub center contracts distances between pairs of vertices within hubs, and between groups of vertices across hubs, yielding a small characteristic path length. In addition, reduced sized neighborhoods, and the inclusion of hub centers into neighborhoods, serve to sustain the clustering coefficient at its initially high value. We find that optimized networks are more clustered than their corresponding regular networks, and have a smaller average degree of separation than their corresponding random graphs, and hence do better than those described by Watts and Strogatz. Further, small worlds that arise due to optimization require less wiring than their WS counterparts. This may be useful in networks where wiring is expensive. However, degree is not constrained, which allows some vertices to dominate in connectivity.

In summary, our work lends support to the idea that a competitive minimization principle may underlie the formation of a small-world network. Also, we observe that hubs could constitute an integral structural component of any small-world network, and that power laws in edge length scale and vertex connectivity may be signatures of this principle in many complex and diverse systems.

Finally, in future work, we will be studying larger networks that were computationally inaccessible to us at present. We are also investigating the application of the small-world architecture in the brain; also, a dynamic model will be considered to understand the emergence of small worlds in social networks [19].

N.M. thanks V. Vinay and Ramesh Hariharan for useful discussions, and Mark Newman for useful comments and suggestions. This work was performed at the Indian Institute of Science, and submitted as part of Ref. [19].

[1] This is stronger than the usual notion of a regular graph wherein all vertices have the same degree.
[2] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[3] It is instructive to note, however, that the graph though random is not equivalent to a random graph in Ref. [4] due to a constraint imposed by the rewiring which ensures that the degree of each vertex is at least $k/2$.
[4] B. Bollobás, *Random Graphs* (Academic Press, New York, 1985).
[5] C. Cherniak, Trends Neurosci. **18**, 522 (1995).
[6] C. Cherniak, J. Neurosci. **14**, 2418 (1994).
[7] R. Kasturirangan, e-print xxx.lanl.gov/cond-mat/9904055.
[8] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[9] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999).
[10] D. J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton University Press, Princeton, 1999).
[11] S. N. Dorogovtsev and J. F. F. Mendes, Europhys. Lett. **50**, 1 (2000).
[12] M. E. J. Newman, e-print xxx.lanl.gov/abs/cond-mat/0001118.
[13] This explains the use of ''random'' to describe this graph. Although the edges are uniformly distributed as in a random graph, all the edges are not randomly distributed (due to the set of edges that form the universal hub).
[14] A. Barrat and M. Weigt, Eur. Phys. J. B **13**, 547 (2000).
[15] Such vertices have $k_v = 1$, and $|E(\Gamma_v)| = 0$, which results in an invalid definition of C_v. Their clustering coefficient can be taken to be either 0 or 1. Note that, rather than 0, a value of 1 would result in the average clustering coefficient being higher than that at the universal hub.
[16] *COCOON '99*, edited by T. Asano, H. Imai, D. T. Lee, S. Nakano, and T. Tokuyama (Springer-Verlag, Heidelberg, 1999), LNCS No. 1672, pp. 1–17.
[17] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[18] R. Matthews, New Sci. **164**, 24 (2000).
[19] N. Mathias, M. Sci. thesis, Department of Computer Science and Automation, Indian Institute of Science, 2000.