Efficient Behavior of Small-World Networks

Vito Latora[1,2] and Massimo Marchiori[3,4]

[1]*Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Batiment 100, 91405 Orsay Cedex, France*
[2]*Department of Physics and Astronomy, University of Catania, and INFN, 95129 Catania, Italy*
[3]*W3C and Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*
[4]*Department of Computer Science, University of Venice, 30172 Venice, Italy*

(Received 25 January 2001; published 17 October 2001)

We introduce the concept of *efficiency* of a network as a measure of how efficiently it exchanges information. By using this simple measure, small-world networks are seen as systems that are both globally and locally efficient. This gives a clear physical meaning to the concept of "small world," and also a precise quantitative analysis of both weighted and unweighted networks. We study neural networks and man-made communication and transportation systems and we show that the underlying general principle of their construction is in fact a small-world principle of high efficiency.

DOI: 10.1103/PhysRevLett.87.198701 PACS numbers: 89.70.+c, 05.90.+m, 87.18.Sn, 89.40.+k

We live in a world of networks. In fact any complex system in nature can be modeled as a network, where vertices are the elements of the system and edges represent the interactions between them. Coupled biological and chemical systems, neural networks, social interacting species, computer networks, or the Internet are only a few such examples [1]. Characterizing the structural properties of the networks is then of fundamental importance to understand the complex dynamics of these systems. A recent paper [2] has shown that the connection topology of some biological and social networks is neither completely regular nor completely random. These networks, there named "small worlds," in analogy with the concept of small-world phenomenon developed 30 years ago in social psychology [3], are in fact highly clustered like regular lattices, yet have small characteristic path lengths like random graphs. The original paper triggered much interest in the study of the properties of small worlds (see Ref. [4] for a recent review). Researchers have focused their attention on different aspects: study of the inset mechanism [5–7], dynamics [8] and spreading of diseases on small worlds [9], and applications to social networks [10,11] and to the Internet [12,13]. In this Letter we introduce the concept of *efficiency* of a network, measuring how efficiently information is exchanged over the network. By using efficiency, small-world networks are seen as systems that are both globally and locally efficient. This formalization gives a clear physical meaning to the concept of small world, and also allows a precise quantitative analysis of unweighted and weighted networks. We study several systems, such as brains, communication, and transportation networks, and show that the underlying general principle of their construction is in fact a small-world principle, provided attention is taken not to ignore an important observational property (closure).

We start by reexamining the original formulation proposed in Ref. [2]. There, a generic graph $\mathbf{G}$ with N vertices and K edges is considered. $\mathbf{G}$ is assumed to be *unweighted,* i.e., edges are all equal, *sparse* [$K \ll N(N-1)/2$], and *connected;* i.e., there exists at least one path connecting any two vertices with a finite number of steps. $\mathbf{G}$ is therefore represented by simply giving the adjacency (or connection) matrix, i.e., the $N \times N$ matrix whose entry a_{ij} is 1 if there is an edge joining vertex i to vertex j and is 0 otherwise. An important quantity of $\mathbf{G}$ is the degree of vertex i, i.e., the number k_i of edges incident with vertex i (the number of neighbors of i). The average value of k_i is $k = 2K/N$. Once $\{a_{ij}\}$ is given it can be used to calculate the matrix of the shortest path lengths d_{ij} between two generic vertices i and j. The fact that $\mathbf{G}$ is assumed to be connected implies that d_{ij} is positive and finite $\forall i \neq j$. In order to quantify the structural properties of $\mathbf{G}$, Ref. [2] proposes to evaluate two different quantities: the characteristic path length L and the clustering coefficient C. L is the average distance between two generic vertices $L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$, and C is a local property defined as $C = \frac{1}{N} \sum_i C_i$. Here C_i is the number of edges existing in $\mathbf{G}_i$, the subgraph of the neighbors of i, divided by the maximum possible number $k_i(k_i - 1)/2$. In [2] a simple method is considered to produce a class of graphs with increasing randomness. The initial graph $\mathbf{G}$ is taken to be a one-dimensional lattice with each vertex connected to its k neighbors and with periodic boundary conditions. By rewiring each edge at random with probability p, $\mathbf{G}$ can be tuned in a continuous way from a regular lattice ($p = 0$) into a random graph ($p = 1$). For the regular lattice we expect $L \sim N/2k$ and a high clustering coefficient $C = 3/4(k-2)/(k-1)$, while for a random graph $L \sim \ln N / \ln(k-1)$ and $C \sim k/N$ [5,14]. Although in the two-limit cases a large C is associated with a large L and vice versa a small C to a small L, the numerical experiment reveals an intermediate regime at small p where the system is highly clustered like regular lattices, yet having small characteristics path lengths like random graphs. This behavior is there called small world and it is found to be a property of some analyzed social and biological networks [2].

Now we propose a more general setup to investigate real networks. We will show that (i) the definition of small-world behavior can be given in terms of a single variable with a physical meaning, the *efficiency* E of the network; (ii) $1/L$ and C can be seen as first approximations of E evaluated, respectively, on a global and a local scale; and (iii) we can drop all the restrictions on the system, such as unweightedness, connectedness, and sparseness.

We represent a real network as a generic *weighted* (and possibly even *nonsparse* and *nonconnected*) graph $\mathbf{G}$. Such a graph needs two matrices to be described: the adjacency matrix $\{a_{ij}\}$, defined the same as for the unweighted graph, and the matrix $\{\ell_{ij}\}$ of physical distances. The number ℓ_{ij} can be the space distance between the two vertices or the strength of their possible interaction: we suppose ℓ_{ij} to be known even if in the graph there is no edge between i and j. For example, ℓ_{ij} can be the geographical distance between stations in transportation systems (in such a case ℓ_{ij} respects the triangle equality, though this is not a necessary assumption), the time taken to exchange a packet of information between routers in the Internet, or the inverse velocity of chemical reactions along a direct connection in a biological system. Of course, in the particular case of an unweighted graph $\ell_{ij} = 1 \ \forall\, i \neq j$. The shortest path length d_{ij} between two generic points i and j is the smallest sum of the physical distances throughout all the possible paths in the graph from i to j. The matrix $\{d_{ij}\}$ is therefore calculated by using the information contained both in matrix $\{a_{ij}\}$ and in matrix $\{\ell_{ij}\}$. We have $d_{ij} \geq \ell_{ij} \ \forall\, i, j$, the equality being valid when there is an edge between i and j. Let us now suppose that the system is parallel, i.e., every vertex sends information concurrently along the network, through its edges. The efficiency ϵ_{ij} in the communication between vertices i and j can then be defined to be inversely proportional to the shortest distance: $\epsilon_{ij} = 1/d_{ij} \ \forall\, i, j$. When there is no path in the graph between i and j, $d_{ij} = +\infty$ and, consistently, $\epsilon_{ij} = 0$. The average *efficiency* of $\mathbf{G}$ can be defined as

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

To normalize E we consider the ideal case $\mathbf{G}_{id}$ in which the graph $\mathbf{G}$ has all the $N(N-1)/2$ possible edges. In such a case the information is propagated in the most efficient way since $d_{ij} = \ell_{ij} \ \forall\, i, j$, and E assumes its maximum value $E(\mathbf{G}_{id}) = \frac{1}{N(N-1)} \sum_{i \neq j \in \mathbf{G}} \frac{1}{l_{ij}}$. The efficiency $E(\mathbf{G})$ considered in the rest of the paper is always divided by $E(\mathbf{G}_{id})$ and therefore $0 \leq E(\mathbf{G}) \leq 1$. Although the equality $E = 1$ is valid when there is an edge between each couple of vertices, real networks can reach a high value of E.

In our formalism, we can define the small-world behavior by using the single measure E to analyze both the local and the global behavior, rather than two different variables L and C. The quantity in Eq. (1) is the *global efficiency* of $\mathbf{G}$ and we therefore refer to it as E_{glob}. Since E is also defined for a disconnected graph we can characterize the local properties of $\mathbf{G}$ by evaluating for each vertex i the efficiency of $\mathbf{G}_i$, the subgraph of the neighbors of i. We define the *local efficiency* as the average efficiency of the local subgraphs, $E_{\text{loc}} = 1/N \sum_{i \in \mathbf{G}} E(\mathbf{G}_i)$. This quantity plays a role similar to the clustering coefficient C. Since $i \notin \mathbf{G}_i$, the local efficiency E_{loc} reveals how much the system is *fault tolerant,* thus it shows how efficient the communication is between the first neighbors of i when i is removed [15]. The definition of small world can now be rephrased and generalized in terms of the information flow: small-world networks have high E_{glob} and E_{loc}, i.e., are very efficient in global and local communication. This definition is valid both for unweighted and for weighted graphs, and can also be applied to disconnected and/or nonsparse graphs.

It is interesting to see the correspondence between our measure and the quantities L and C of [2] (or, correspondingly, $1/L$ and C). The fundamental difference is that E_{glob} is the efficiency of a *parallel system,* where all the nodes in the network concurrently exchange packets of information (such as all the systems in [2], for example), while $1/L$ measures the efficiency of a *sequential system* (i.e., only one packet of information goes along the network). $1/L$ is a reasonable approximation of E_{glob} when there are no huge differences among the distances in the graph, and this can explain why L works reasonably well in the unweighted examples of [2]. But, in general, $1/L$ can significantly depart from E_{glob}. For instance, in the Internet, having a few computers with extremely slow connections does not mean that the entire Internet efficiency is diminished: in practice, the presence of such very slow computers goes unnoticed, because the other thousands of computers are exchanging packets among them in a very efficient way. Here $1/L$ would give a number very close to zero (strictly 0 in the particular case when a computer is disconnected from the others and $L = +\infty$), while E_{glob} gives the correct efficiency measure of the Internet. We now turn our attention to the local properties of a network. C is only one among the many possible intuitive measures [10] of how well connected a cluster is. It can be shown that, when in a graph, most of its local subgraphs $\mathbf{G}_i$ are not sparse, then C is a good approximation of E_{loc}. In summary, there are *not* two different types of analyses to be done for the global and local scales, just one with a very precise physical meaning: the efficiency in transporting information.

We now illustrate the onset of the small world in an unweighted graph by means of the same example used in [2]. A regular lattice with $N = 1000$ and $k = 20$ is rewired with probability p, and E_{glob} and E_{loc} are reported in Fig. 1 as functions of p [16]. For $p = 0$ we expect the system to be inefficient on a global scale $[E_{\text{glob}} \sim k/N \log(N/K)]$ but locally efficient. The situation is inverted for the random graph. In fact, at $p = 1$, E_{glob} assumes a maximum value of 0.4, meaning 40% efficiency of the ideal graph with an edge between each pair

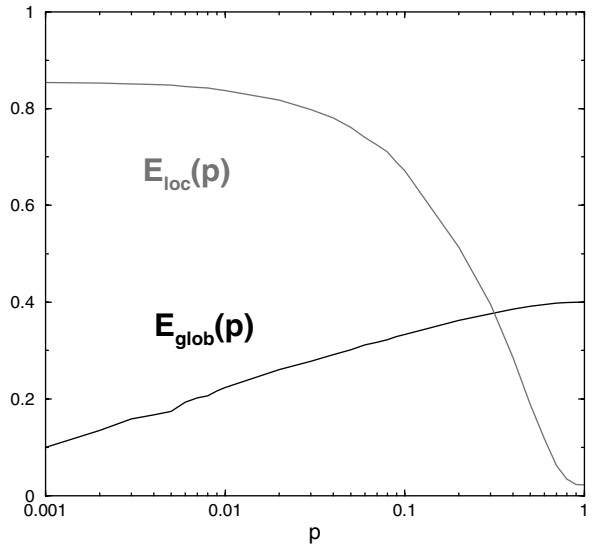

FIG. 1. Global and local efficiency for the graph example considered in [2]. A regular lattice with $N = 1000$ and $k = 20$ is rewired with probability p. The small-world behavior results from the increase of E_{glob} caused by the introduction of only a few rewired edges (short cuts), which on the other side do not affect E_{loc}. At $p \sim 0.1$, E_{glob} has almost reached the value of the random graph, though E_{loc} has diminished only by very little from the value of 0.82 of the regular lattice. Small worlds have high E_{glob} and E_{loc}.

of vertices. This occurs at the expense of the fault tolerance ($E_{\text{loc}} \sim 0$). The small-world behavior appears for intermediate values of p. It results from the fast increase of E_{glob} (for small p we find a linear increase of E_{glob} in the logarithmic horizontal scale) caused by the introduction of only a few rewired edges (short cuts), which on the other side do not affect E_{loc}. At $p \sim 0.1$, E_{glob} has almost reached the maximum value of 0.4, though E_{loc} has diminished only very little from the maximum value of 0.82. For an unweighted case the description in terms of network efficiency resembles the approximation given in [2]. In particular we have checked that a good agreement with curves $L(p)$ and $C(p)$ [2] can be obtained by reporting $1/E_{\text{glob}}(p)$ and $E_{\text{loc}}(p)$. Of course in such an example the short cuts connect at almost no cost vertices that would otherwise be much farther apart (because $\ell_{ij} = 1$ $\forall\, i \neq j$). On the other hand, this is not true when we consider a weighted network. As real networks, we first consider different examples of natural systems (neural networks), and then we turn our attention to man-made communication and transportation systems.

(a) Neural networks.—Thanks to recent experiments, neural structures can be studied at several levels of scale. Here we focus first on the analysis of the neuroanatomical structure of the cerebral cortex, and then on a simple nervous system at the level of wiring between neurons. The anatomical connections between cortical areas are of particular importance for their intricate relationship with the functional connectivity of the cerebral cortex [18]. We analyze two databases of cortico-cortical connections in the macaque and in the cat [19]. Table I indicates that the two networks are small worlds [16]: they have high E_{glob}, 52% and 69% of the efficiency of the ideal graph with an edge between each pair of vertices (just slightly smaller than the best possible values of 57% and 70% obtained in random graphs), and high E_{loc}, 70% and 83%, i.e., high fault tolerance [22]. These results indicate that in the neural cortex each region is intermingled with the others and grows following a perfect balance between local necessities (fault tolerance) and wide-scope interactions. Next we consider the neural network of C. elegans, the only case of a nervous system completely mapped at the level of neurons and chemical synapses [23]. Table I shows that this is also a small-world network: C. elegans achieves 50% of both global and local efficiency. Moreover the value of E_{glob} is similar to E_{loc}. This is a difference from cortex databases, where fault tolerance is slighty privileged with respect to global communication.

TABLE I. Macaque and cat cortico-cortical connections [19]. The macaque database contains $N = 69$ cortical areas and $K = 413$ connections [20]. The cat database has $N = 55$ cortical areas (including hippocampus, amygdala, entorhinal cortex, and subiculum) and $K = 564$ (revised database and cortical parcellation from [21]). The nervous system of C. elegans consists of $N = 282$ neurons and $K = 2462$ links which can be either synaptic connections or gap junctions [24].

	E_{glob}	E_{loc}
Macaque	0.52	0.70
Cat	0.69	0.83
C. elegans	0.46	0.47

(b) Communication networks.—We have considered two of the most important large-scale communication networks present today: the World Wide Web (WWW) and the Internet. Table II shows that they have relatively high values of E_{glob} (slightly smaller than the best possible values obtained for random graphs) and E_{loc}. Despite the fact that the WWW is a virtual network and the Internet is a physical network, at a global scale they transport information essentially in the same way (as their E_{glob}'s are almost equal). At a local scale, the bigger E_{loc} in the WWW case can be explained both by the tendency in the WWW to create Web communities (where pages talking about the same subject tend to link to each other) and by the fact that many pages within the same site are often quickly connected to each other by some root or menu page.

(c) Transport networks.—Differently from previous databases, the Boston subway transportation system

TABLE II. Communication networks. Data on the World Wide Web from http://www.nd.edu/~networks contains $N = 325\,729$ documents and $K = 1\,090\,108$ links [12], while the Internet database is taken from http://moat.nlanr.net and has $N = 6474$ nodes and $K = 12\,572$ links.

	E_{glob}	E_{loc}
WWW	0.28	0.36
Internet	0.29	0.26

(MBTA) can be better described by a weighted graph, the matrix $\{\ell_{ij}\}$ being given by the geographical distances between stations. If we consider the MBTA as an unweighted graph we find that it is apparently neither locally nor globally efficient (see Table III). On the other hand, when we take the geographical distances into account, we obtain $E_{\text{glob}} = 0.63$: this shows the MBTA is a very efficient transportation system on a global scale, only 37% less efficient than the ideal subway with a direct tunnel from each station to the others. Even in the weighted case E_{loc} stays low (0.03), indicating poor local behavior: differently from a neural network, the MBTA is not fault tolerant and damage in a station will dramatically affect the connection between the previous and the next station. The difference with respect to neural networks comes from different needs and priorities in the construction and evolution mechanism: when we build a subway system, the priority is given to the achievement of global efficiency, and not to fault tolerance. In fact a temporary problem in a station can be solved by other means: for example, walking or taking a bus from the previous to the next station. That is to say, the MBTA is not a *closed system*. It can be considered, after all, a subgraph of a wider transportation network. This property is of fundamental importance when we analyze a system: while global efficiency is without doubt the major characteristic, it is *closure* that somehow leads a system to have high local efficiency (without alternatives, there should be high fault tolerance). The MBTA is not a closed system, and thus this explains why, unlike in the case of the brain, fault tolerance is not a critical issue. Indeed, if we increase the precision of the analysis and change the MBTA subway network by taking into account, for example, the Boston bus system, this extended transportation system comes back to be a small-world network ($E_{\text{glob}} = 0.72$, $E_{\text{loc}} = 0.46$). Qualitatively similar results, confirming the similarity of construction principles, have been obtained for other undergrounds and for a wider transportation system consisting of all the main airplane and highway connections throughout the world [25]. Considering all the transportation alternatives available at that scale makes again the system closed (there are no other reasonable routing alternatives), and so fault tolerance comes back as a leading construction principle.

In summary, the introduction of the efficiency measure allows us to give a definition of small world with a clear physical meaning, and provides important hints on why the original formulas of Ref. [2] work reasonably well in some cases and where they fail. The efficiency measure allows a precise quantitative analysis of the information flow, and works both in the unweighted abstraction and in the more realistic assumption of weighted networks. Finally, analysis of real data indicates that various existing (neural, communication, and transport) networks exhibit small-world behavior (even, in some cases, when their unweighted abstractions do not), substantiating the idea that the diffusion of small-world networks can be interpreted as the need to create networks that are both globally and locally efficient.

TABLE III. The Boston underground transportation system (MBTA) consists of $N = 124$ stations and $K = 124$ tunnels. The matrix $\{\ell_{ij}\}$ of the spatial distances between stations, used for the weighted case, has been calculated using databases from http://www.mbta.com/ and the U.S. National Mapping Division.

	E_{glob}	E_{loc}
MBTA (unweighted)	0.10	0.006
MBTA (weighted)	0.63	0.03

[1] Y. Bar-Yam, *Dynamics of Complex Systems* (Addison-Wesley, Reading, MA, 1997).
[2] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[3] S. Milgram, Psychol. Today **2**, 60 (1967).
[4] M. E. J. Newman, cond-mat/0001118.
[5] A. Barrat and M. Weigt, Eur. Phys. J. B **13**, 547 (2000).
[6] M. Marchiori and V. Latora, Physica (Amsterdam) **285A**, 539 (2000).
[7] M. Barthelemy and L. Amaral, Phys. Rev. Lett. **82**, 3180 (1999).
[8] L. F. Lago-Fernandez *et al.,* Phys. Rev. Lett. **84**, 2758 (2000).
[9] C. Moore and M. E. J. Newman, Phys. Rev. E **61**, 5678 (2000).
[10] M. E. J. Newman, cond-mat/0011144.
[11] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[12] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[13] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[14] B. Bollobás, *Random Graphs* (Academic, London, 1985).
[15] Our concept of fault tolerance is different from the one adopted in R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000) and R. Cohen *et al.,* Phys. Rev. Lett. **85**, 2758 (2000), where the authors consider the response of the entire network to the removal of i.
[16] Here, and in the following, the matrix $\{d_{ij}\}_{i,j\in G}$ was computed by using two different methods: the Floyd-Warshall $[O(N^3)]$ [17] and the Dijkstra algorithm $[O(N^2 \log N)]$ [10].
[17] G. Gallo and S. Pallottino, Ann. Oper. Res. **13**, 3 (1988).
[18] O. Sporns, G. Tononi, and G. M. Edelman, Cerebral Cortex **10**, 127 (2000).
[19] J. W. Scannell, Nature (London) **386**, 452 (1997).
[20] M. P. Young, Philos. Trans. R. Soc. London B **252**, 13 (1993).
[21] J. W. Scannell, M. P. Young, and C. Blakemore, J. Neurosci. **15**, 1463 (1995).
[22] E. Sivan, H. Parnas, and D. Dolev, Biol. Cybern. **81**, 11 (1999).
[23] J. G. White *et al.,* Philos. Trans. R. Soc. London B **314**, 1 (1986).
[24] T. B. Achacoso and W. S. Yamamoto, *AY's Neuroanatomy of C. elegans for Computation* (CRC, Boca Raton, FL, 1992).
[25] M. Marchiori and V. Latora (to be published).