Harmony in the small-world

Massimo Marchiori[a,b], Vito Latora[c,*]

[a]*The World Wide Web Consortium (W3C), USA*
[b]*MIT Laboratory for Computer Science, USA*
[c]*Dipartimento di Fisica, Universitá di Catania and INFN Sezione, Corso Italia 57, 95129 Catania, Italy*

Abstract

The small-world phenomenon, popularly known as six degrees of separation, has been mathematically formalized by Watts and Strogatz in a study of the topological properties of a network. Small-world networks are defined in terms of two quantities: they have a high clustering coefficient C like regular lattices and a short characteristic path length L typical of random networks. Physical distances are of fundamental importance in applications to real cases; nevertheless, this basic ingredient is missing in the original formulation. Here, we introduce a new concept, the *connectivity length D*, that gives harmony to the whole theory. D can be evaluated on a global and on a local scale and plays in turn the role of L and $1/C$. Moreover, it can be computed for any metrical network and not only for the topological cases. D has a precise meaning in terms of information propagation and describes in a unified way, both the structural and the dynamical aspects of a network: small-worlds are defined by a small global and local D, i.e., by a high efficiency in propagating information both on a local and global scale. The neural system of the nematode *C. elegans*, the collaboration graph of film actors, and the oldest US subway system, can now be studied also as metrical networks and are shown to be small-worlds. © 2000 Elsevier Science B.V. All rights reserved.

PACS: 05.90.+m; 87.18.Sn; 87.23.G; 89.40.+k

Keywords: Neural networks; Social systems; Small world

Many biological and social systems present in nature, such as neural networks or nervous systems of living organisms, structured societies, but also Internet and the World Wide Web, can be represented by means of a graph. In such a graph, the vertices are the single units (the elements) of the system (neurons in the brain, human beings in a society, single computer in a computer network, etc.), while the edges

---

* Corresponding author. Fax: +39-095-383023.
E-mail address: latora@ct.infn.it (V. Latora).

are the links or a measure of the interactions between the single elements. If we are interested in understanding the complex dynamical behaviour of many natural systems it is very important to study the topological and the metrical properties of the underlying networks.

The small-world behaviour [1,2], popularly known as six degrees of separation [3], has been mathematically formalized by Watts and Strogatz in a study of topological networks [4,5]. The original formulation of small-world abstracts from the real distances in a network: it only deals with the topology of the connections, and imposes further constrains on the graph structure, like total connectedness [6].

In this paper, we propose a general theory of small-world networks. We start by considering a generic metrical graph G, that in principle, can be also non-connected. N is the number of vertices (or nodes) in the graph and K is the total number of edges (or arcs). Two nodes i and j of the graph are at a certain *physical distance* $\ell_{i,j}$, which can be for example the real distance between the two nodes or a measure of the strength of their possible interaction. The *distance on the graph* $d_{i,j}$ is instead defined by the shortest sum of the physical distances throughout all the possible paths in the graph from i to j. Let us suppose that every node sends information along the network, through its arcs. And, every node in the network propagates information concurrently. The amount of information sent from the node i to the node j per unit of time is $v/d_{i,j}$, where v is the velocity at which the information travels over the network. When there is no path in the network between i and j, we assume $d_{i,j} = +\infty$, and consistently, the amount of exchanged information is 0. The *performance* of G is the total amount of information propagated per unit of time over the network:

$$P = \sum_{i,j \in G} v/d_{i,j} \, . \tag{1}$$

Every sum here and in the following is intended for $i \neq j$. If we want to quantify the typical separation between two vertices in the graph, a good measure is given by the *connectivity length* $D(G)$: this is the fixed distance to which we have to set every two vertices in the graph in order to maintain its performance. Interestingly enough, the connectivity length of the graph is *not the arithmetic mean* but the *harmonic mean* of all the distances:

$$D(G) = H(\{d_{i,j}\}_{i,j \in G}) = \frac{N(N-1)}{\sum_{i,j \in G} 1/d_{i,j}} \, . \tag{2}$$

The harmonic mean has been known since the time of Pythagoras and Plato [7] as the mean expressing "harmonious and tuneful ratios", and later has been employed by musicians to formalize the diatonic scale, and by architects [8] as a paradigm for beautiful proportions [9]. Nowadays, it finds extensive applications in a variety of different fields, like traffic [10], information retrieval [11], visibility systems [12], water control [13] and many others. In particular, the harmonic mean is used to calculate the average performance of computer systems [14,15], parallel processors [16], and communication devices (for example, modems and Ethernets [17]). In all such cases,

where a mean flow-rate of information has to be computed, the simple arithmetic mean gives the wrong result.

The definition of small-world proposed by Watts and Strogatz is based on two different quantities: a measure of the global properties of the graph (L), defined as the average number of edges in the shortest path between two vertices, and a local quantity (C), measuring the average cliquishness properties of a generic vertex. C is the average number of edges existing in the clique of a generic vertex (the graph of its neighbors) divided by the maximum possible number of edges in the clique. The main reason to introduce C is because L, defined as the simple arithmetic mean of $d_{i,j}$, applies only to connected graphs and cannot be used for cliques subgraphs, that in most of the cases are disconnected. In our theory, we have a uniform description both of the global and of the local properties of the network by means of the single measure D. In fact, we can evaluate D_{glob}, the connectivity length for the global graph G, and D_{loc} the average connectivity length of its cliques. Small-world networks are defined by small D both on global and on local scale. The connectivity length D gives harmony to the whole theory of small-world networks, since:

- It is not just a generic intuitive notion of average distance in a network, but has a precise meaning in terms of network *efficiency*.
- It describes in an unified way the system on a global and on a local scale.
- It applies both to topological and *metrical* networks. The topological approximation [4] is too strong an abstraction for any real network. Here we show that small-worlds are indeed present in nature, and are not just a topological effect.
- It applies to any graph, not only to connected graphs as the original theory [4].
- It describes the structural, but also the *dynamical* aspects of a network. In the second part of their article, Watts and Strogatz try to investigate the dynamics of a small-world network by means of an example model of disease spreading [4]. Using numerical simulations, they find out that the time of disease propagation and L have a similar form (see their Fig. 3(b)). In our theory this result and the very same concept of effectiveness in the dynamics of signal propagation in a small-world network are already implicit in the definition of D: small D_{glob} and D_{loc} mean a high performance both on a local and on a global scale.

We now present a few numerical examples and some applications to real networks. As a first numerical experiment, we consider the original topological example used by Watts and Strogatz [4]. In Fig. 1 a small-world graph is constructed from a regular lattice with $N = 1000$ vertices and 10 edges per vertex, by means a random rewiring process that introduces increasing disorder with probability p. Such a procedure does not consider the geometry of the system: in fact the physical distance $\ell_{i,j}$ between any two vertices is always equal to 1. Here and in the rest of the paper, we calculate the distances on the graph $\{d_{i,j}\}_{i,j \in G}$ by means of the Floyd–Warshall algorithm [18]. This method is extremely efficient and allows to compute in parallel all the distances in the graph. Moreover, D is normalized to its minimum value $H(\{\ell_{i,j}\}_{i,j \in G})$ and ranges in $[1, +\infty]$. A similar normalization is also used in Ref. [4]. In panels (a) and (b) we plot the global and the local connectivity length versus p. The rewiring process
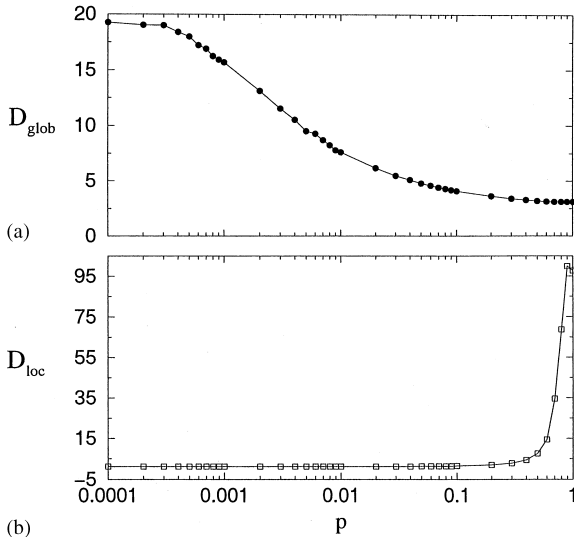
Fig. 1. Topological case. We consider $N = 1000$, 10 edges per vertex and the same random rewiring used by Watts and Strogatz [4]. In panels (a) and (b) we plot the global and the local connectivity length versus p. The data shown in the figure are averages over 10 random realizations of the rewiring process. The logarithmic horizontal scale is used to resolve the rapid drop in D_{glob} due to the presence of short cuts and corresponding to the onset of the small-world.

produces small-world networks which result from the immediate drop in D_{glob} caused by the introduction of a few long-range edges. Such "short cuts" connect vertices that would otherwise be much farther apart at no cost because $\ell_{i,j} = 1 \; \forall i \neq j$. During the drop of D_{glob}, D_{loc} remains small and almost equal to the value for the regular lattice. Small-worlds are characterized by having small D_{glob} and D_{loc}, i.e., by high global and local performances.

We can now compare our definition with the original one given by Watts and Strogatz. In Fig. 2 we report the same quantities of Fig. 1 normalized by $D_{glob}(0)$ and $D_{loc}(0)$, i.e., the values for the regular lattice. Our figure reproduces perfectly the results shown in Fig. 2 of Ref. [4]. With this normalization, the global and local connectivity lengths behave, respectively, like L and $1/C$, and in the topological case our theory gives the same results of Watts and Strogatz.

The essential role of short cuts is emphasized by a second experiment in which we repeat the test by changing the metric of the system. In Fig. 3 we implement a random rewiring in which the length of each rewired edge is set to change from 1 to 3. This time short cuts have a cost. The figure shows that the small-world behaviour is still present even when the length of the rewired edges is larger than the original one and the behaviour of D_{glob} is not just monotonically decreasing. To check the robustness of these results we increased even more the length of the rewired edges, and we have tested many different metrical networks (points on a circle or on a bidimensional square lattice), obtaining similar results.
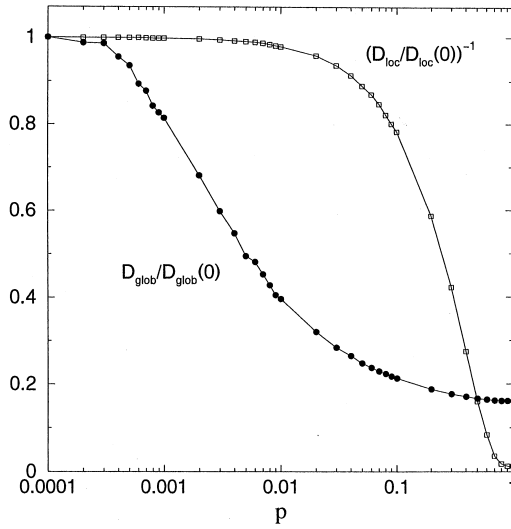
Fig. 2. We normalize D_{glob} and D_{loc}, respectively, by $D_{\text{glob}}(0)$ and $D_{\text{loc}}(0)$ and we reproduce Fig. 2 of Watts and Strogatz [4]. L is equivalent to D_{glob}, while C plays the same role as $1/D_{\text{loc}}$.
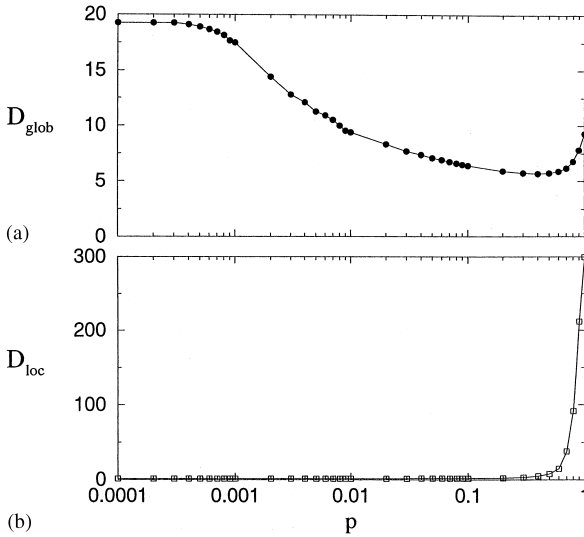

Fig. 3. Metrical case. We consider $N = 1000$, 10 edges per vertex and we implement a random rewiring in which the length of each rewired edge is set to change from 1 to 3. The data shown are averages over 10 random realizations of the rewiring process.

This second numeric example suggests that the small-world behaviour is not only an effect of the topological abstraction but can be found in nature in all such cases, where the physical distance is important and the rewiring has a cost. Therefore, we have studied three real networks of great interest:

Table 1
C. elegans[a]

	D_{glob}	$D_{\mathrm{glob}}^{\mathrm{random}}$	D_{loc}	$D_{\mathrm{loc}}^{\mathrm{random}}$
Topological	2.19	2.15	2.12	12.69
Metrical	31.51	31.28	8.64	36.93

[a]Connectivity lengths of the real network compared to random graphs show that the *C. elegans* is both a topological and a metrical small-world.

Table 2
Film actors[a]

	D_{glob}	$D_{\mathrm{glob}}^{\mathrm{random}}$	D_{loc}	$D_{\mathrm{loc}}^{\mathrm{random}}$
Topological	3.37	3.01	1.87	1800
Metrical	47.2	58.6	37.6	2600

[a]Same as in Table 1 for the collaboration graph of actors. Both the topological and the metrical graph show the small-world phenomenon.

- The neural network of the nematode worm *C. elegans* [19], the only case of completely mapped neural network existing in the market.
- The collaboration graph of actors in feature films [20], which is an example of a non-connected social network.
- The Massachusetts Bay underground transportation system [21], a case in which $\{\ell_{i,j}\}$ are given by the space distances between stations i and j.

These cases are interesting because they are all better described by metrical rather than topological graphs.

In Table 1 we report the results on the *C. elegans*. The *C. elegans* nervous system consists of $N = 282$ neurons and two different types of links, synaptic connections and gap junctions, through which information is propagated from neuron to neuron. In the topological case studied by Watts and Strogatz, the graph consists of $K = 2462$ edges, each defined, when two vertices are connected by either a synapse or a gap junction [4]. This is only a crude approximation of the real network. Neurons are different from one another and some of them are in much stricter relation than others: the number of junctions between two neurons can vary a lot (up to a maximum of 72). A metrical network is more suited to describe such a system and can be defined by setting $\ell_{i,j}$ equal to the inverse number of junctions between i and j. Connectivity lengths of real networks, compared to random graphs, show that the *C. elegans* is both a topological and a metrical small-world.

In Table 2 we study the collaboration graph of actors extracted from the Internet Movie Database [20], as of July 1999. The graph has $N = 277336$ and $K = 8721428$, and is not a connected graph. The approach of Watts and Strogatz cannot be applied directly and they have to restrict their analysis to the giant connected component of the graph [4]. Here we apply our small-world analysis directly to the whole graph, without any restriction. Moreover, the topological case only provides whether actors participated in some movie together, or if they did not at all. Of course, in reality there are instead

Table 3
T Boston[a]

	D_{glob}	D_{glob}^{random}	D_{loc}	D_{loc}^{random}
Metrical	1.58	11.06	25.87	250

[a]Connectivity lengths of the real network, compared to random graphs show that the Massachusetts Bay transportation system is a metrical small-world.

various degrees of correlation: two actors who have done 10 movies together are in a much stricter relation rather than two actors who have acted together only once. As in the case of *C. elegans* we can better shape this different degree of friendship by using a metrical network: we set the distance $\ell_{i,j}$ between two actors i and j as the inverse of the number of movies they did together. The numerical values in Table 2 indicate that both the topological and the metrical network show the small-world phenomenon.

The Massachusetts Bay transportation system [21] (MBTA, popularly known as *T*) is the oldest subway system in the US (the first electric streetcar line in Boston, which is now part of the MBTA Green Line, began its operation on January 1, 1889) and consists of $N = 124$ stations and $K = 124$ tunnels extending through Boston and the other cities of the Massachusetts Bay. This is an example of an important real network, where the matrix $\ell_{i,j}$ is given by the spatial distances between two stations i and j [22]. We have calculated such distances using information databases from Geographic Data Technology (GDT), the US Defence Mapping Agency, and the National Mapping Division. The comparison with random graphs in Table 3 indicates the MBTA is a small-world network, thus is a very efficient transportation system. D also gives a precise quantitative information: $D_{glob} = 1.58$ shows that the MBTA is only 58% less efficient than the optimal subway ($D = 1$) with a direct connection tunnel for each couple of stations. On the other side, the relatively high value of D_{loc} indicates that the system is not perfectly *fault tolerant* [23]. This is intuitively explained by the fact that usually most of the network is blocked if a tunnel in the subway is interrupted.

To conclude, the values of D_{glob} and D_{loc} in Tables 1–3 show that all the above three cases (topological and metrical) are small-worlds. These real examples, coming from different fields indicate that the small-world phenomenon is not merely an artifact of an oversimplified topological model but a common characteristic of biological and social systems. The theory presented here provides with new general tool of investigation for any complex system in nature and it can be applied to a huge number of real cases. Furthermore, it incorporates in a unified way both the structural and the dynamical aspects of a network.

References

[1] S. Milgram, The small world problem, Physicol. Today 2 (1967) 60–67.
[2] M. Kochen, The Small World, Ablex, Norwood, NJ, 1989.
[3] J. Guare, Six Degrees of Separation: A Play, Vintage Books, New York, 1990.
[4] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, Nature 393 (1998) 440–442.
[5] J.J. Collins, C.C. Chow, It's a small world, Nature 393 (1998) 409–410.
[6] B. Bollabás, Random Graphs, Academic, London, 1985.
[7] Plato, Timaeus, 347 B.C.
[8] R. Wittkower, Architectural Principles in the Age of Humanism, 1949.
[9] A. Palladio, The Four Books of Architecture, 1530 (reprinted by Dover, New York, 1976).
[10] D.L. Gerlough, M.J. Huber, Traffic Flow Theory: A Monograph (U.S. Transportation Research Board Special Report, 1975).
[11] C.J. van Rijsbergen, Information Retrieval, Butterworth, London, 1979.
[12] U.S. National Weather Service Training Center, The Automated Surface Observing System (ASOS) Algorithm Tutorial.
[13] U.S. Environmental Protection Agency, Technical Support Document for Water Quality-based Toxics Control, EPA 440/4-85-032, April 1990.
[14] J. Smith, Characterising computer performance with a single number, Commun. ACM 31 (10) (1988) 1202–1206.
[15] M.L. Van Name, B. Catchings, Benchmark testing: advancing the art, PC Mag. 1421 (1996) 13–17.
[16] K. Hwang, F.A. Briggs, Computer Architecture and Parallel Processing, McGraw-Hill, New York, 1988.
[17] R. Jain, The Art of Computer Systems Performance Analysis, Wiley, New York, 1991.
[18] G. Gallo, S. Pallottino, Ann. Oper. Res. 13 (1988) 3–79.
[19] T.B. Achacoso, W.S. Yamamoto, AY's Neuroanatomy of C. Elegans for Computation, CRC Press, Boca Raton, FL, 1992.
[20] The Internet Movie Database, http://www.imdb.com
[21] Massachusetts Bay Transportation Authority, http://www.mbta.com/
[22] M. Marchiori, V. Latora, Small-world behaviour in Spatio-Temporal Systems, in preparation.
[23] E. Sivan, H. Parnas, D. Dolev, Fault tolerance in the cardiac ganglion of the lobster., Biol. Cybernet. 81 (1999) 11–23.