

Communicability Angle and the Spatial Efficiency of Networks*

Ernesto Estrada[†]
Naomichi Hatano[‡]

Abstract. We introduce the concept of communicability angle between a pair of nodes in a graph. We provide strong analytical and empirical evidence that the average communicability angle for a given network accounts for its spatial efficiency on the basis of the communications among the nodes in a network. We determine characteristics of the spatial efficiency of more than a hundred real-world networks that represent complex systems arising in a diverse set of scenarios. In particular, we find that the communicability angle correlates very well with the experimentally measured value of the relative packing efficiency of proteins that are represented as residue networks. We finally show how we can modulate the spatial efficiency of a network by tuning the weights of the edges of the networks. This allows us to predict the effects of external stresses on the spatial efficiency of a network as well as to design strategies to improve important parameters in real-world systems.

Key words. complex network, communicability, graph distance, graph planarity, Euclidean distance

AMS subject classifications. 05C12, 05C50, 05C82, 05C10

DOI. 10.1137/141000555

I. Introduction. Networks are ubiquitous in many real-world scenarios, ranging from the biomolecular to the social and infrastructural organization of modern society [11, 32, 9]. In many of these networks, nodes and edges are used to represent physically embedded objects [4], namely, *spatial networks*. Some examples of spatial networks are urban street networks [25], brain and other anatomical networks [7], protein residue networks [11], infrastructural networks [4], the networks of channels in fractured rocks, the networks representing the corridors and galleries in animal nests, and many more; for even more, see [11] and references therein.

A natural question that arises in the analysis of spatial networks is how efficiently they use the available geometrical space in which they are embedded. The concept of spatial efficiency accounts for many organizational principles of complex systems, ranging from the folding of proteins into globular shapes in order to minimize the volume occupied inside the cell [10], to the planarity of urban street networks [8] that poses a serious challenge to their continuous growth in view of their threat to the natural environment.

The concept of *spatial efficiency* is adapted here from economics, where it is frequently used to describe how much time, effort, and cost a given arrangement

*Received by the editors December 22, 2014; accepted for publication (in revised form) November 17, 2015; published electronically November 3, 2016.

<http://www.siam.org/journals/sirev/58-4/100055.html>

[†]Department of Mathematics & Statistics, University of Strathclyde, Glasgow G11XQ, UK (ernesto.estrada@strath.ac.uk).

[‡]Institute of Industrial Science, University of Tokyo, Komaba, Meguro, Tokyo 153-8505, Japan (hatano@iis.u-tokyo.ac.jp).

produces for governments, businesses, and households when conducting their activities as compared to alternative arrangements; see [35] and references therein. This concept has a lot to do with the efficiency in communication among the parts of the system under consideration, and as such it presents a well-posed problem for its analysis beyond spatial networks. Indices for communication efficiency of networks have been previously proposed in the literature [27, 28, 1, 21], and they revolve around the idea of considering the sum of reciprocal shortest-path distances in graphs [27]. Here we extend this concept beyond the shortest path by means of the *communicability function* [15], which quantifies how much information can flow from one node to another in a network; see also [16, 17].

In the present article we cover two very important areas of network theory, namely, the relationship between network structure and communication efficiency as well as the inference of spatial properties of networks based on their topological properties. Our theoretical constructions are based on a remarkable mapping of each node of a network to a point on the surface of a hypersphere. This embedding is obtained using concepts from matrix functions, making a nice bridge between linear algebra, geometry, and network theory. In fact, we prove here that the distance defined on the basis of the communicability function is indeed the chord distance between the two points on the hypersphere. We can thereby assign a Euclidean angle to each pair of nodes which represents the communication efficiency between them. We show here that this *communicability angle* accounts for the *spatial efficiency* of networks. This connection between matrix functions and spatial properties of networks allows us to bridge areas of network theory that have so far remained separate. For instance, we show some remarkable connections between the communicability angle and the graph planarity which may represent new avenues connecting algebraic and topological graph theory. Last but not least, this paper develops a few applications from different areas which should be of interest to applied mathematicians working in mathematical social sciences, mathematical biology/ecology, and the growing arena of city data analytics.

2. Preliminaries. In this section we shall present some of the definitions, notations, and properties associated with networks in order to make this work self-contained. Here we consider networks represented by *graphs* $\Gamma = (V, E)$ having n nodes (vertices) and m edges (links). The reader is directed to [11] for the standard notation used here. We designate by k_p the degree of the node p . A *walk* of length ℓ in Γ is a set of nodes $p_1, p_2, \dots, p_\ell, p_{\ell+1}$ such that for all $1 \leq i \leq \ell$, $(p_i, p_{i+1}) \in E$. A *closed walk* is a walk for which $p_1 = p_{\ell+1}$. A *path* is a walk with no repeated nodes. Throughout this work, we will always consider undirected, simple, and connected networks. The following standard notation will be used through this paper: P_n represents the path graph; K_n represents the complete graph; and K_{n_1, n_2} represents the complete bipartite graph; in particular, $K_{1, n-1}$ is known as the star graph. We recall that a graph is planar if it can be drawn on a plane without any edge crossings [23].

We let A denote the adjacency matrix, which for undirected simple finite graphs is a real symmetric matrix. We can decompose it into the form

$$(2.1) \quad A = U \Lambda U^T,$$

where Λ is a diagonal matrix containing the eigenvalues of A , which we label in nonincreasing order $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, and $U = [\vec{\psi}_1, \dots, \vec{\psi}_n]$ is an orthogonal matrix, where $\vec{\psi}_\mu$ is an eigenvector associated with λ_μ . Because we consider connected graphs, A is irreducible; the Perron–Frobenius theorem then dictates that $\lambda_1 > \lambda_2$ and that we can choose $\vec{\psi}_1$ such that its components $\psi_1(p)$ are positive for all $p \in V$.

An important quantity for studying communication processes in networks is the communicability function [15, 17, 16], defined for each pair of nodes p and q as

$$(2.2) \quad G_{pq} = \sum_{k=0}^{\infty} \frac{(A^k)_{pq}}{k!} = (e^A)_{pq} = \sum_{\mu=1}^n e^{\lambda_{\mu}} \psi_{\mu}(p) \psi_{\mu}(q).$$

The factor $(A^k)_{pq}$ counts the number of walks of length k starting at the node p and ending at the node q . The communicability function is the sum of the numbers of walks of length k , each weighted by the factor $1/k!$ so that shorter walks may be more influential than longer ones. In (2.2), the exponential of the matrix A is defined by its Taylor expansion, which is the communicability function itself. The spectral decomposition on the right-hand side is also derived from the spectral decomposition of each term of the Taylor expansion: $(A^k)_{pq} = \sum_{\mu} (\lambda_{\mu})^k \psi_{\mu}(p) \psi_{\mu}(q)$.

The importance of the communicability function (2.2) lies in the fact that it takes account of long walks as well as short ones; even two nodes connected by a very long shortest path can have a strong communication if they are connected by many longer walks. The diagonal term G_{pp} characterizes the degree of participation of the node p in all subgraphs of the network. It is thus known as the subgraph centrality of the corresponding node [18].

We can visualize the communicability function (2.2) in another way. Consider a matrix-vector equation $d\vec{\psi}/dt = A\vec{\psi}$, which governs the time evolution of a vector $\vec{\psi}(t)$. If the vector $\vec{\psi}(t)$ describes a random-walker distribution on the network in question at time t , the above equation describes how the walkers move around on the network. Its formal solution is given by $\vec{\psi}(t) = e^{At}\vec{\psi}(0)$, and hence the exponential matrix e^A is the time evolution operator for the unit time. Therefore, the communicability function (2.2) is the transition rate for the walkers on the site p (represented by a vector \vec{w}_p) to move to the site q (represented by another vector \vec{w}_q) after unit time, where \vec{w}_p is a column vector with unity on the p th element and zero on the others.

Other quantities used in this work, such as the network efficiency E [27], the resistance distance Ω_{pq} [26], the Cheeger constant $i(G)$, and the modularity index Q are defined in the supplementary information accompanying this paper. In the next section we will introduce another distance measure recently defined on the basis of the communicability function. It is novel in the sense that walks longer than the shortest path are taken into account.

3. Communicability Distance. The new distance function is defined as [12, 13]

$$(3.1) \quad \xi_{pq}^2 = G_{pp} + G_{qq} - 2G_{pq},$$

which we will refer to as the communicability distance between the nodes p and q in Γ . The intuition behind it is that when two nodes p and q communicate with each other, the quality of their communication depends on two factors: (i) how much information departing from the node p (q) arrives at the node q (p), and (ii) how much information departing from the node p (q) returns to that node p (q) without arriving at its destination. That is, the communication efficiency increases with the amount of information that departs from the originator and arrives at its destination, but decreases with the amount of information that is frustrated due to the fact that the information returns to its originator without being delivered to its target. We can rephrase the information flow as random walkers according to the interpretation that e^A is a time-evolution operator for unit time. This intuition has led to the definition (3.1).

It has indeed been proved that the function ξ_{pq} is a Euclidean distance between the nodes p and q in Γ [12].

THEOREM 3.1 (see [19]). *The communicability distance ξ_{pq} induces an embedding of the graph Γ of size n into a hypersphere of radius $R^2 = [c - (2 - b)^2/a]/4$ in an $(n - 1)$ -dimensional space, where $a = \vec{1}^T e^{-A} \vec{1}$, $b = \vec{s}^T e^{-A} \vec{1}$, and $c = \vec{s}^T e^{-A} \vec{s}$ with $\vec{s} = \text{diag } e^A$.*

Let us now give a more intuitive and geometric view of the communicability distance. For this purpose, we first prove the following theorem.

THEOREM 3.2. *Let $\vec{x}_p = e^{A/2} \vec{\phi}_p$, where $\vec{\phi}_p = (\psi_1(p) \ \cdots \ \psi_\mu(p) \ \cdots \ \psi_n(p))^T$. Then we have*

$$(3.2) \quad G_{pq} = \vec{x}_p \cdot \vec{x}_q.$$

(The proof is given in the supplementary information.)

This theorem transforms the communicability distance (3.1) into the form

$$(3.3) \quad \xi_{pq}^2 = \vec{x}_p \cdot \vec{x}_p + \vec{x}_q \cdot \vec{x}_q - 2\vec{x}_p \cdot \vec{x}_q = |\vec{x}_p - \vec{x}_q|^2.$$

In other words, the communicability distance is the Euclidean distance in the space of $\{\vec{x}_p\}$. In order to visualize this space, let us go back to the interpretation that $(e^A)_{pq}$ is the transition rate of the random walkers from the p th site to the q th site. An initial state \vec{w}_p is a basis vector in the original vector space, but its expression is given by $\vec{\phi}_p$ above in the vector space with the eigenvectors $\vec{\psi}_\mu$ as its basis vectors, namely, the eigenspace; see Figure 1 for an example for the path graph P_3 . In other words, the initial vector represents the state in which all random walkers sit on the p th site, but it is denoted by the vector $\vec{\phi}_p$ in the eigenspace. The vector \vec{x}_p is a vector in the eigenspace, representing the state in which random walkers from the p th site move around for time $1/2$.

Theorem 3.1 dictates that the vectors $\{\vec{x}_p\}$ fall onto the surface of a hypersphere in the space; see Figure 2(a) for an illustration in the case $n = 3$. We can understand this in the following way. We first fix the n -dimensional normal vector \vec{x}_\perp from n pieces of conditions $(\vec{x}_p - \vec{x}_\perp) \cdot \vec{x}_\perp = 0$ for $1 \leq p \leq n$. This specifies the $(n - 1)$ -dimensional flat surface on which all vectors fall as $(\vec{x} - \vec{x}_\perp) \cdot \vec{x}_\perp = 0$. We next fix the n -dimensional vector \vec{x}_0 that specifies the center of the hypersphere as well as the radius R from $n + 1$ pieces of conditions $(\vec{x}_0 - \vec{x}_\perp) \cdot \vec{x}_\perp = 0$ and $|\vec{x}_p - \vec{x}_0| = R$ for $1 \leq p \leq n$.

We can therefore regard ξ_{pq} as the chord distance between the two points on the hypersurface. Figure 2(b) shows the triangle spanned by the vectors \vec{x}_p and \vec{x}_q . This leads to the definition in the next section of the angle between the two vectors.

4. Communicability Angle. Let p and q be nodes of a connected simple network and let us define the following quantity:

$$(4.1) \quad \gamma_{pq} := \frac{G_{pq}}{\sqrt{G_{pp}G_{qq}}}.$$

We then prove the following result.

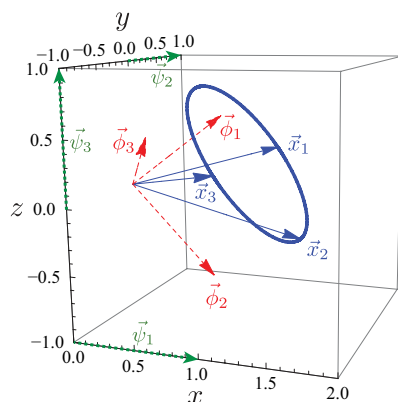


Fig. 1 A demonstration plot of $\vec{\phi}_p$ (red dashed arrows) and \vec{x}_p (blue solid arrows) for the path graph P_3 . The eigenvectors $\vec{\psi}_\mu$ (green dotted arrows) define the axes of this eigenspace. The communicability distance ξ_{pq} is the chord distance on the (blue) circle that goes through the end points of the vectors \vec{x}_p .

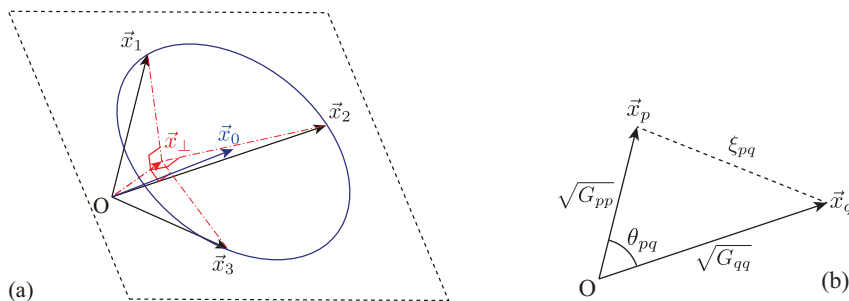


Fig. 2 (a) Three vectors \vec{x}_1 , \vec{x}_2 , and \vec{x}_3 (solid black arrows) in a three-dimensional space spanned by the three eigenvectors of a 3×3 adjacency matrix A . The vectors fall on a two-dimensional flat surface (broken black lines) to which the vector \vec{x}_\perp (red dot-dashed arrow) is normal. We can draw a circle (solid blue curve) on the two-dimensional surface around a point \vec{x}_0 (solid blue arrow) to contain all three points. (b) The triangle spanned by the vectors \vec{x}_p and \vec{x}_q .

THEOREM 4.1. The index γ_{pq} is the cosine of the Euclidean angle spanned by the position vectors of p and q .

(The proof is given in the supplementary information.)

We then call θ_{pq} the communicability angle between the corresponding nodes of the graph. Details on how to compute the communicability angle for networks are given in the Supplementary Information accompanying the present paper. For each pair of nodes in the graph, the communicability distance and angle are related mathematically by the following expression:

$$(4.2) \quad \xi_{pq}^2 = G_{pp} + G_{qq} - 2\sqrt{G_{pp}G_{qq}} \cos \theta_{pq}.$$

Because $G_{pq} \geq 0$ for any pair of nodes in Γ , the communicability angle is bounded by $0 \leq \cos \theta_{pq} \leq 1$, that is, the communicability angle of simple graphs can take values only in the range $(0^\circ, 90^\circ)$. We will now describe some classes of graphs that show how we attain the extremal values.

PROPOSITION 4.2. Let P_n be the path graph with n nodes labeled by $1, 2, \dots, n$ sequentially. The communicability angle between any pair of nodes in P_n is given by

$$(4.3) \quad \cos \theta_{pq}(P_n) = \frac{I_{p-q}(2) - I_{p+q}(2)}{\sqrt{[I_0(2) - I_{2r(p)}(2)][I_0(2) - I_{2r(q)}(2)]}}$$

in the limit $n \rightarrow \infty$, where $I_\gamma(z)$ is the Bessel function of the first kind and

$$(4.4) \quad r(p) = \begin{cases} p & \text{for } p \leq n/2 \text{ with even } n \text{ or } p \leq (n+1)/2 \text{ with odd } n, \\ n-p+1 & \text{for } p > n/2 \text{ with even } n \text{ or } p > (n+1)/2 \text{ with odd } n. \end{cases}$$

(The proof is given in the supplementary information.)

Notice that for the pair of nodes at the ends of the path we have

$$(4.5) \quad \lim_{n \rightarrow \infty} \cos \theta_{n1}(P_n) = \lim_{n \rightarrow \infty} \frac{I_{n-1}(2) - I_{n+1}(2)}{I_0(2) - I_2(2)} = 0,$$

which attains the upper bound of the communicability angle.

PROPOSITION 4.3. Let $K_{1,n-1}$ be the star graph with n nodes. Let the node with degree $n-1$ be labeled as 1. The communicability angle between any pair of nodes in $K_{1,n-1}$ is given by

$$(4.6) \quad \cos \theta_{1q}(K_{1,n-1}) = \frac{\tanh^2(\sqrt{n-1})}{(n-2) \operatorname{sech}(\sqrt{n-1}) + 1} \quad \text{for } q \neq 1,$$

$$(4.7) \quad \cos \theta_{pq}(K_{1,n-1}) = \frac{\cosh(\sqrt{n-1}) - 1}{(n-2) \cosh(\sqrt{n-1}) + n - 2} \quad \text{for } p \neq 1 \text{ and } q \neq 1.$$

(The proof is given in the supplementary information.)

It is important to notice that

$$(4.8) \quad \lim_{n \rightarrow \infty} \cos \theta_{1q}(K_{1,n-1}) = 1 \quad \text{for } q \neq 1,$$

$$(4.9) \quad \lim_{n \rightarrow \infty} \cos \theta_{pq}(K_{1,n-1}) = 1 \quad \text{for } p \neq 1 \text{ and } q \neq 1,$$

which attain the lower bound of the communicability angle.

PROPOSITION 4.4. Let K_n be the complete graph with n nodes. The communicability angle between any pair of nodes in K_n is given by

$$(4.10) \quad \cos \theta_{pq} = \frac{e^n - 1}{e^n + n - 1}.$$

(The proof is given in the supplementary information.)

Notice that $\cos \theta_{pq} \rightarrow 1$ as $n \rightarrow \infty$ in K_n .

In the supplementary information we provide some analytic results that allow us to understand the main differences between the communicability distance and the communicability angle. Let us present the structural information provided by the results there, which will be useful for further application of the communicability angle in analyzing real-world networks. Let us define the average communicability angle for a given graph as the average over the pairs of nodes:

$$(4.11) \quad \langle \theta \rangle = \frac{2}{n(n-1)} \sum_{p>q} \theta_{pq}.$$

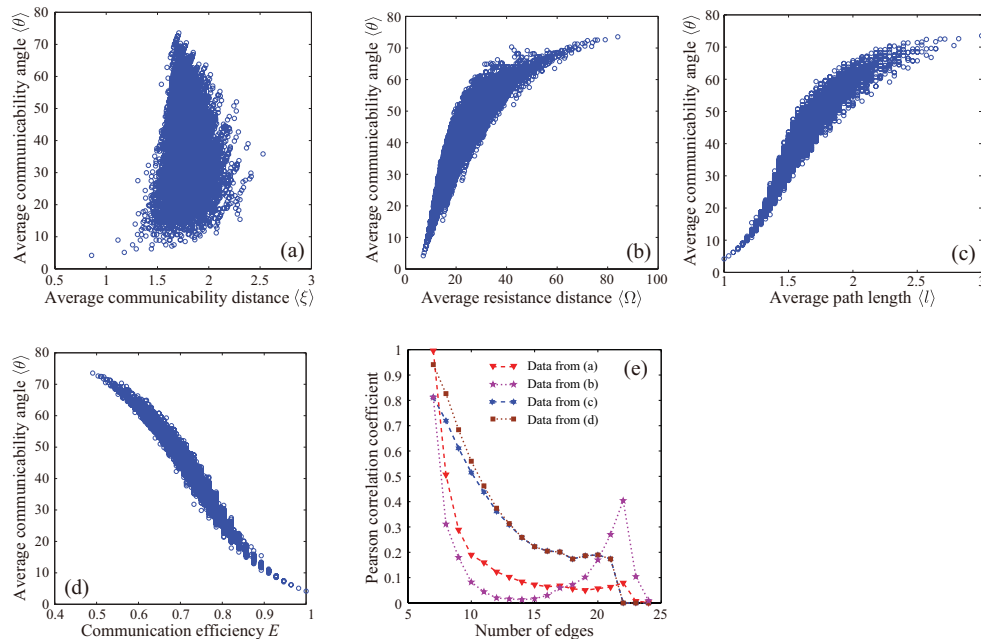


Fig. 3 (a)–(d) Scatter plots of the average communicability angle against (a) the average communicability distance, (b) the average resistance distance, (c) the average path length, and (d) the communication efficiency for all 11,117 connected graphs with 8 nodes. (e) The squared Pearson correlation coefficients between the metrics for 8-node graphs with a fixed number of edges.

We then have the following observations: (i) The average communicability angle for the path graph P_n tends to 90° when the number of nodes tends to infinite. This is a consequence of Propositions 4.2 and SI3.2 (the latter presented in the supplementary information). (ii) The average communicability angle for the star graph $K_{1,n-1}$ tends to 0° when the number of nodes tends to infinite. This is a consequence of Proposition 4.3. (iii) The average communicability angle for the complete graph K_n tends to 0° when the number of nodes tends to infinite. This is a consequence of Proposition 4.4.

5. Computational Analysis of the Communicability Angle. In this section we computationally analyze the average communicability angle $\langle \theta \rangle$ in (4.11) for connected graphs. Specifically, we study a dataset of all 11,117 connected graphs with 8 nodes. We divide this section into three subsections: we first analyze relationships (or lack thereof) between the average communicability angle and other graph metrics, namely, the average path length, the average resistance distance, and the average communicability distance; we then study relationships between $\langle \theta \rangle$ and the graph planarity; we finally investigate the influence of graph modularity on the communicability angle.

5.1. Communicability Angle and Other Graph Metrics. We first compare the average communicability angle $\langle \theta \rangle$ with the average communicability distance $\langle \xi \rangle$, the average resistance distance $\langle \Omega \rangle$, the average path length $\langle l \rangle$, and the communication efficiency E as metrics potentially related to $\langle \theta \rangle$; each average was taken over all pairs of nodes. We show in Figure 3 the scatter plots of these measures against the average communicability angle.

We can see that the communicability angle is neither directly nor trivially related to the other metrics. It is particularly interesting to see the lack of correlation between $\langle\theta\rangle$ and $\langle\xi\rangle$, which are highly uncorrelated although the two quantities are based on the same concept of communicability. This lack of correlation is not unexpected if we consider how the two measures and the communicability function are related to each other via (4.2). The average communicability angle shows more similar trends to the average path length $\langle l \rangle$, the average resistance distance $\langle \Omega \rangle$, and the communication efficiency E . The extreme values of these three measures coincide with those of $\langle\theta\rangle$, although there is a large dispersion inbetween. The general plots in Figure 3 really hide the true lack of correlation that exists among these metrics and the communicability angle. To reveal the lack of correlation more explicitly, we plot the squared Pearson correlation coefficient between each metric and $\langle\theta\rangle$ for groups of graphs having the same number of edges. As can be seen in Figure 3(e), as soon as the number of edges increases, the correlation between the pair of indices drops significantly. For instance, let us consider the communication efficiency, for which the correlation with $\langle\theta\rangle$ yields a correlation coefficient $r^2 \approx 0.94$ for the 8-node trees. This correlation coefficient drops to $r^2 \approx 0.31$ for graphs having 13 edges and to $r^2 \approx 0.17$ for graphs having 18 edges. It is virtually zero for graphs with more than 22 edges. The reason for this decay in the correlation is very important. Trees have very large correlations between the pairs of measures. This is due to the fact that in these graphs there are only shortest paths to connect any pair of nodes because of the absence of any cycles. As the number of edges increases, the number of potential routes between any pair of nodes increases dramatically, making the measures based on shortest paths more different from the communicability angle. There is also a complete lack of correlation between the communicability angle and the average resistance distance for graphs having 10 to 20 edges. The correlation coefficient increases for these two measures when the number of edges is 23, but then decays. The reason for this increase is not clear at all, but the peak height is only 0.4 anyway.

Among all the connected graphs with 8 nodes, the path graph P_8 has the largest average communicability angle and the complete graph K_8 has the smallest. Among all the trees with 8 nodes, the star graph $K_{1,7}$ has the smallest average communicability angle. This is also verified for all connected graphs with 5, 6, and 7 nodes. We thereby have the following conjectures.

CONJECTURE 5.1. *Among all connected graphs with n nodes, the average communicability angle is the largest for the path graph P_n and the smallest for the complete graph K_n .*

CONJECTURE 5.2. *Among all trees with n nodes, the average communicability angle is the largest for the path graph P_n and the smallest for the star graph $K_{1,n-1}$.*

These observations indicate that the average communicability angle describes the efficiency of a graph in using the space in which it is embedded. The path graph P_n , which intuitively occupies the largest portion of space, has the largest average communicability angle, while the star and complete graphs, which intuitively occupy the smallest, have the average communicability angle close to zero. In the next section we explore more observations of this sort from a computational point of view.

5.2. Communicability Angle and Graph Planarity. Here we investigate the relationship between the graph planarity and the average communicability angle. We first determine whether or not a graph is planar using the planarity test proposed

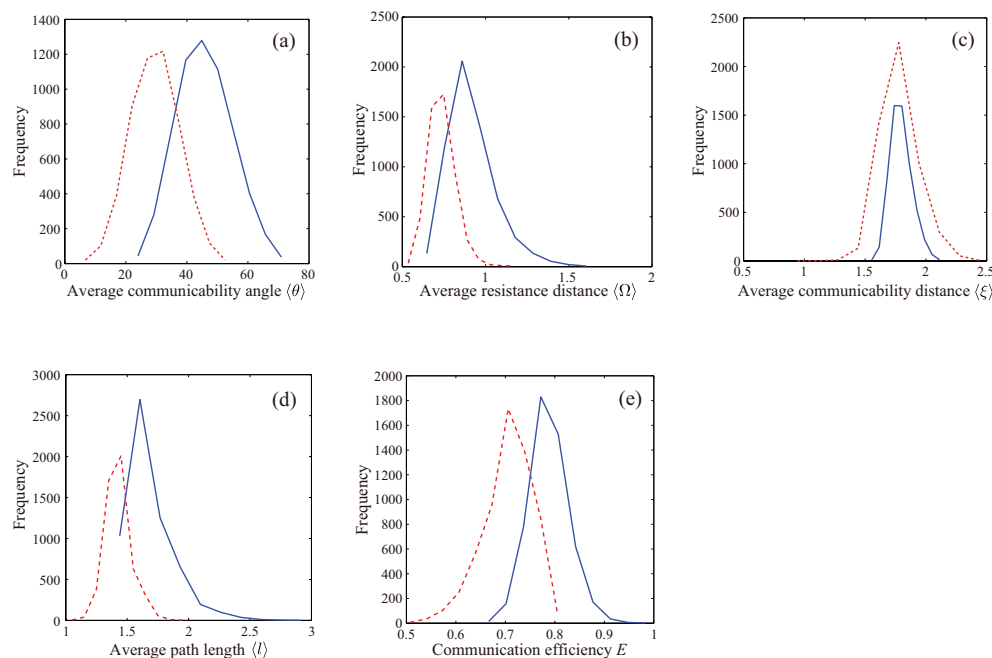


Fig. 4 Frequency of planar and nonplanar graphs for different values of (a) the average communicability angle, (b) the average resistance distance, (c) the average communicability distance, (d) the average path length, and (e) the communication efficiency. The histogram for planar graphs is shown as a solid line and that for nonplanar graphs as a broken line.

by Boyer and Myrvold [6]. We then construct the histogram of the frequency of planar/nonplanar graphs with respect to the average communicability angle.

Let η_k be the number of planar graphs having $k \leq \langle\theta\rangle < (k + 10^\circ)$ for $k = 0^\circ, 10^\circ, 20^\circ, \dots, 80^\circ$. We plot in Figure 4(a) the histogram of the planar/nonplanar graphs as a function of their values of $\langle\theta\rangle$ for all connected graphs with 8 nodes. For comparison, we also show similar plots in Figure 4(b)–(d) for the average resistance distance $\langle\Omega\rangle$, the communicability distance $\langle\xi\rangle$, and the average path length $\langle l\rangle$.

The first interesting observation is that the planar graphs yield significantly larger values of $\langle\theta\rangle$ than the nonplanar graphs. The peaks in the histogram Figure 4(a) for the planar and nonplanar graphs are at $\langle\theta\rangle \approx 44.87^\circ$ and $\langle\theta\rangle \approx 32.17^\circ$, respectively. There is a larger relative separation between the two peaks of the histogram for $\langle\theta\rangle$ than for the rest of the measures. Let us put this in a quantitative context. Let us define the percentage of variation between the maxima of the two peaks as $v(\%) = 100 \times (x_h(\text{planar}) - x_h(\text{nonplanar})) / (x_{\max} - x_{\min})$, where $x_h(\dots)$ is the value of the corresponding variable for the peak in the histogram, while x_{\max} and x_{\min} are the maximum and minimum values, respectively, of the variable x for the whole dataset of 8-node graphs. For instance, for $x = \langle\theta\rangle$, the values are $x_h(\text{planar}) = 44.87^\circ$, $x_h(\text{nonplanar}) = 32.17^\circ$, $x_{\max} = 73.55^\circ$, and $x_{\min} = 4.19^\circ$. Then the percentages of the variation between the maxima of the two peaks are 18.3% for $\langle\theta\rangle$, 13.4% for E , 9.7% for $\langle\Omega\rangle$, and 7.7% for $\langle l\rangle$. As is obvious from Figure 4(c), this percentage is zero for the communicability distance. We have repeated these experiments by considering all the 261,080 connected graphs with 9 nodes, and the results are as follows: 23.2% for $\langle\theta\rangle$, 15.3% for E , 10.7% for $\langle\Omega\rangle$, and 9.4% for $\langle l\rangle$. Thus, it is clear

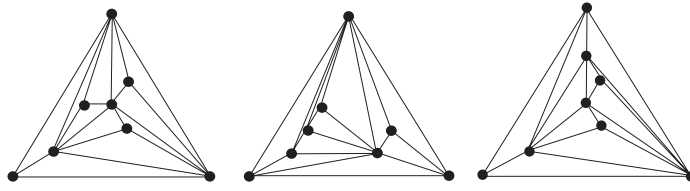


Fig. 5 Three maximal planar graphs with 8 nodes which have the smallest values of $\langle \theta \rangle$. The graphs are drawn as triangulations using Schnyder embedding [36]. Because the graphs are maximal planar, adding any edge will make the resulting graph nonplanar.

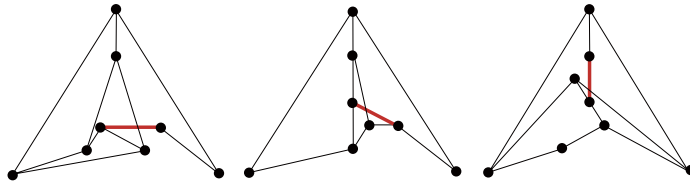


Fig. 6 Three minimal nonplanar graphs with 8 nodes which have the largest values of $\langle \theta \rangle$. The graphs are drawn using the Schnyder embedding [36] and allowing the superposition of one edge (marked in red thick line) whose removal will transform the graph into a planar one.

that the communicability angle not only shows the best separation between planar and nonplanar graphs, but also has the largest increase in this separation when increasing the number of nodes.

We can elaborate more on the relationship between planarity and the communicability angle from the analysis of the connected graphs with 8 nodes: (i) No planar graph has $\langle \theta \rangle < 21.4^\circ$. (ii) The planar graphs with the smallest value of $\langle \theta \rangle$ correspond to the *maximal planar graphs*. A graph is maximal planar, also known as a *triangulation*, if the addition of any edge to it results in a nonplanar graph. Obviously, these are the “least planar” of all planar graphs. Examples are given in Figure 5. (iii) There is no nonplanar graph with $\langle \theta \rangle > 55.065^\circ$. (iv) The nonplanar graphs with the largest values of $\langle \theta \rangle$ are minimal nonplanar graphs. A minimal nonplanar graph is a nonplanar graph for which every proper subgraph is planar, i.e., removing any node or edge makes the graph planar. Again, these are the “least nonplanar” of all the nonplanar graphs. Examples are given in Figure 6.

The previous results do not necessarily mean that the average communicability angle characterizes the graph planarity, or vice versa, but instead that the planarity is indeed an important ingredient of the spatial efficiency as measured by the communicability angle.

5.3. Communicability Angle and Graph Modularity. Modularity is a very important concept for the study of real-world networks. It refers to the property of graphs with clusters of highly interconnected nodes but with poor intercluster connectivity. Such clusters are usually referred to as communities in network theory and are expected to play fundamental organizational roles in real-world networks, e.g., groups of proteins with similar actions and groups of people with common interests.

As a first example we construct random modular graphs in the following manner. We generate random modular graphs with 1000 nodes and 50 modules. Then with a fixed total edge density we systematically increase the proportion of edges within

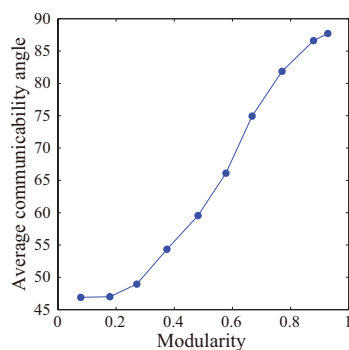


Fig. 7 Relationship between the Newman modularity index [33] and the average communicability angle for random modular graphs with 1000 nodes and 50 modules. The total edge density is 0.01 and the proportion of intra- to intermodular edges varies from 0.1 to 0.95. The points in the plot indicate the average of 100 random realizations. The line is to guide the eye.

modules compared to edges across modules. As this proportion of intramodular edges to intermodular edges increases, the graphs become more modular in the sense previously explained. In order to capture the degree of modularity of these graphs we use the Newman modularity index [33], which is defined as

$$(5.1) \quad Q = \sum_{k=1}^{n_C} \left[\frac{E_k}{m} - \frac{1}{4m^2} \left(\sum_{j \in V_k} k_j \right)^2 \right],$$

where E_k is the number of edges in the k th module, n_C is the total number of modules, m the total number of edges, and k_j the node degree.

In Figure 7 we illustrate the results of plotting the modularity of the random modular graphs and the average communicability angle. As can be seen, as the modularity tends to its maximum, the average communicability angle tends to 90° , indicating the decrease in the spatial efficiency of these graphs.

A network with such clusters has structural bottlenecks; that is, if small groups of nodes/edges are removed, the network is disconnected into two or more relatively large connected components. An extreme case is that of the dumbbell graphs K_n-K_n , that is, two cliques of n nodes connected by only one edge; the removal of the edge separates the network into two connected components of $n/2$ nodes each.

On the other hand, a superhomogeneous graph, which is usually referred to as a good expansion graph, is characterized by the fact that every subset S with more than $n/2$ nodes has a large boundary, which is the number of edges with one node inside the set S and the other in \bar{S} [34]. Expander graphs are characterized by having a large spectral gap $\lambda_1 - \lambda_2$ of the adjacency matrix [2]; see [24, 29] for details.

What is important for the present subsection is that expanders are characterized by their lack of modularity, i.e., the lack of tightly connected clusters which are poorly interconnected by structural bottlenecks. In networks where $\lambda_1 \gg \lambda_2$, we have the following expression for the communicability angle:

$$(5.2) \quad \cos \theta_{pq} = \frac{G_{pq}}{\sqrt{G_{pp}G_{qq}}} \simeq \frac{\psi_1(p) \psi_1(q) e^{\lambda_1}}{\sqrt{\psi_1(p)^2 e^{\lambda_1} \psi_1(q)^2 e^{\lambda_1}}} = \cos 0^\circ.$$

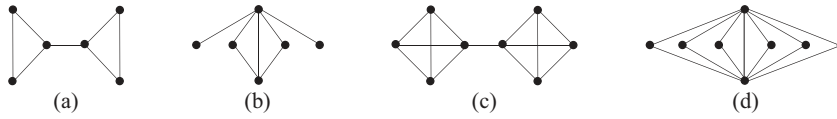


Fig. 8 The graphs with 6 nodes and 7 edges with (a) the largest and (b) the smallest average communicability angles. The same holds for the graphs with 8 nodes and 13 edges in (c) and (d).

That is, the networks lacking any modularity are characterized by a very small value of the communicability angle. On the other hand, in a network where λ_1 is not significantly larger than λ_2 , we make use of the expansions

$$G_{pp}G_{qq} = \psi_1(p)^2\psi_1(q)^2e^{2\lambda_1} + (\psi_1(p)^2\psi_2(q)^2 + \psi_2(p)^2\psi_1(q)^2)e^{\lambda_1+\lambda_2} + \psi_2(p)^2\psi_2(q)^2e^{2\lambda_2} + \text{h.o.}, \quad (5.3)$$

$$G_{pq}^2 \simeq \psi_1(p)^2\psi_1(q)^2e^{2\lambda_1} + 2\psi_1(p)\psi_1(q)\psi_2(p)\psi_2(q)e^{\lambda_1+\lambda_2} + \psi_2(p)^2\psi_2(q)^2e^{2\lambda_2} + \text{h.o.}, \quad (5.4)$$

where h.o. denotes the higher-order terms. The communicability angle is thereby transformed into the form

$$\cos \theta_{pq} = \frac{G_{pq}}{\sqrt{G_{pq}^2 + (\psi_1(p)\psi_2(q) - \psi_2(p)\psi_1(q))^2 e^{\lambda_1+\lambda_2} + \text{h.o.}}}. \quad (5.5)$$

The second term in the denominator depends on the size of the spectral gap; the closer λ_2 is to λ_1 , i.e., the smaller the spectral gap, the larger the denominator is, and consequently, the smaller (5.5) is. Therefore, the angle θ_{pq} becomes larger as the spectral gap grows smaller. We should remark here that θ_{pq} does not depend only on the spectral gap because the higher-order terms in (5.5) can make an important contribution.

Let us give some examples that illustrate the above important relationship between the communicability angle and the graph modularity. Here we again focus on $\langle \theta \rangle$. We first consider the dumbbell graph K_3-K_3 shown in Figure 8(a). It consists of two cliques of 3 nodes each which are connected by a link, thus having 7 edges in total. The average communicability angle for this graph is $\langle \theta \rangle \approx 57.105$ and its spectral gap is $\Delta \approx 0.682$. Among the 19 graphs with 6 nodes and 7 edges, the dumbbell K_3-K_3 has the largest value of $\langle \theta \rangle$. The smallest value of the average communicability angle is obtained for the graph in Figure 8(b), having $\langle \theta \rangle \approx 47.935$ and $\Delta \approx 2.284$.

The situation is very similar for the 1,454 graphs with 8 nodes and 13 edges, among which the dumbbell graph K_4-K_4 in Figure 8(c) has the largest average communicability angle $\langle \theta \rangle \approx 53.876$ with the spectral gap $\Delta \approx 0.511$. The graph with the smallest value of $\langle \theta \rangle$ is the so-called agave graph shown in Figure 8(d); it consists of 2 connected nodes, each of which is also connected to the other $n - 2$ nodes that are not connected among themselves. It has $\Delta = 4.00$ and $\langle \theta \rangle \approx 31.782$. The graphs with the second and third smallest average communicability angles, $\langle \theta \rangle \approx 35.123$ and $\langle \theta \rangle \approx 35.606$ with $\Delta \approx 2.988$ and $\Delta \approx 3.337$, respectively, have structures similar to the agave graph. Notice that the agave graph can be disconnected by removing two edges, but the remaining principal connected component has $n - 1$ nodes, while the removal of 50% of the edges in this graph creates a principal connected component

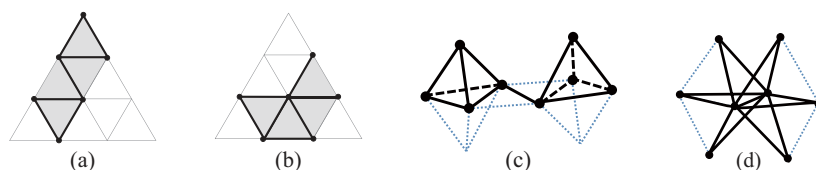


Fig. 9 (a)–(b) Planar embeddings of the graphs in Figure 8(a)–(b), respectively, onto triangular lattices. (c)–(d) Three-dimensional embeddings of the graphs in Figure 8(c)–(d), respectively, onto close-packed lattices.

still containing 62.5% of the nodes. This shows the robustness of this graph to edge removal, a characteristic of good expander graphs due to the lack of any structural bottleneck.

Figure 9(a)–(b) shows planar embeddings of the graphs in Figure 8(a)–(b), respectively, onto triangular lattices. The shadowed areas indicate the triangles covered by the graphs in these embeddings. Although both cover the four triangles, the latter graph, the one with the smallest average communicability angle, covers the most efficient packing in two-dimensional space, which is the area with a node surrounded by six others forming a hexagon. This is known as the penny-packing problem; see [22] for further information. The embedding of the graph with higher modularity and the largest average communicability angle is far from this optimal configuration.

A similar situation occurs with the graphs in Figure 8(c)–(d), the ones with the largest and smallest $\langle \theta \rangle$ among those with 8 nodes and 13 edges; Figure 9(c)–(d) show their embeddings onto close-packed lattices. We can conclude from these observations that a large average communicability angle indicates a poor spatial efficiency of the graph, while a small value of $\langle \theta \rangle$ is associated with the efficient use of space.

5.4. Communicability Angle and Graph Holes. Another characteristic of spatial efficiency that is desirable to be captured by the average communicability angle is the existence of holes. The presence of large holes in a graph obviously makes its spatial efficiency very poor. For instance, let us consider a city in which all the streets form an annulus such that the whole center of the city is empty. The density of streets in that city is very small in comparison to what it is expected from the area occupied by the whole city.

Here we propose to consider the Sierpinski graphs as a model of simple graphs embedded in a Euclidean space such that the density of the graph decays with the size. By the density we mean here the number of nodes divided by the area occupied by the corresponding external triangle. Let us denote by

$$(5.6) \quad \vec{e}_1 = (1, 0, 0), \quad \vec{e}_2 = (0, 1, 0), \quad \vec{e}_3 = (0, 0, 1)$$

the canonical basis vectors of \mathbb{R}^3 . The Sierpinski graphs are generated iteratively from $G_0 = (V_0, E_0)$, where $V_0 = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ and $E_0 = \{(\vec{e}_1, \vec{e}_2), (\vec{e}_2, \vec{e}_3), (\vec{e}_3, \vec{e}_1)\}$. Then, for $G_k = (V_k, E_k)$ we have [37]

$$(5.7) \quad V_{k>0} = (2^{k-1}\vec{e}_1 + V_{k-1}) \cup (2^{k-1}\vec{e}_2 + V_{k-1}) \cup (2^{k-1}\vec{e}_3 + V_{k-1}),$$

$$(5.8) \quad E_{k>0} = (2^{k-1}\vec{e}_1 + E_{k-1}) \uplus (2^{k-1}\vec{e}_2 + E_{k-1}) \uplus (2^{k-1}\vec{e}_3 + E_{k-1}),$$

where \uplus represents the disjoint union of sets. We illustrate in Figure 10 the Sierpinski graphs G_1 , G_2 , and G_3 . The total area occupied by the graph is the area of the

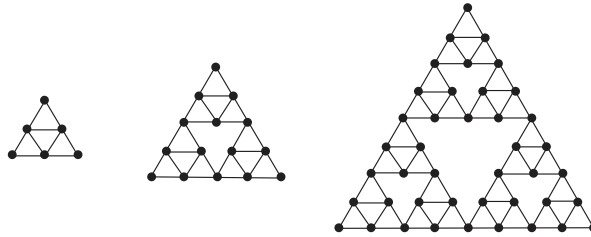


Fig. 10 The Sierpinski graphs G_1, G_2 , and G_3 (from left to right).

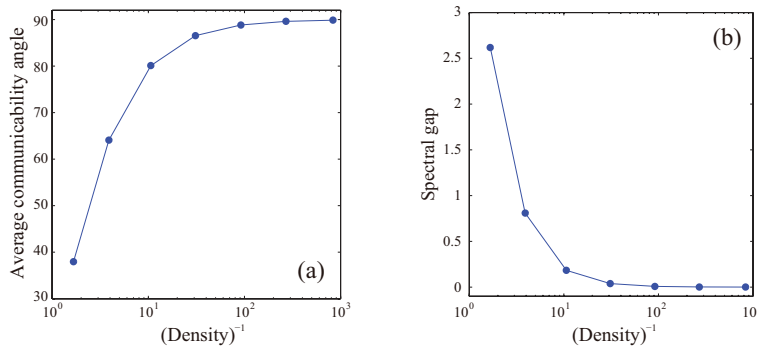


Fig. 11 Plots of (a) the variation of the average communicability angle and (b) the spectral gap of the adjacency matrix with the reciprocal of the density of the Sierpinski graphs. The lines here are to guide the eye.

external triangle which has coordinates $(2k, 0, 0)$, $(0, 2k, 0)$, $(0, 0, 2k)$. Notice that the Sierpinski graphs G_0 and G_1 do not have any holes, G_2 has a central hole of length 6, and G_3 has a central hole of length 12 plus three holes of length 6. As the graph grows, G_k has a central hole of length $2^{k-1} \times 3$ with more holes of smaller sizes, and hence becomes more “spongy.”

We have created the Sierpinski graphs for $k = 1, \dots, 7$ and calculated their densities, defined as the number of nodes divided by the area of the external triangle. We illustrate in Figure 11(a) the relationship between the reciprocal of the density of the Sierpinski graphs and the average communicability angle. For G_1 , which contains no hole, the communicability angle is $\langle \theta \rangle \approx 37.96^\circ$ although the graph is planar. As the size of the graphs increases the average communicability angle quickly goes to its maximum for simple graphs, e.g., $\langle \theta \rangle \approx 90^\circ$ for G_7 , which has 3,282 nodes. The results illustrated in Figure 11(a) agree with our intuition that the communicability angle accounts for the spatial efficiency of graphs. A Sierpinski graph with a large number of nodes containing very large holes (e.g., the graph G_7 has a central hole of length 192, as well as many other holes of smaller sizes) lacks spatial efficiency in the sense of not using appropriately all the available space covered by the external triangles. In order to understand mathematically this relationship we need to use the concept of isoperimetric number (SI-2). We recall that a graph with a small isoperimetric number contains structural holes and/or bottlenecks, which are indications of poor spatial efficiency. Thus, we should expect that a large Sierpinski graph has a very small isoperimetric constant. Mohar [30] has found the following spectral bounds

for the isoperimetric number of a graph:

$$(5.9) \quad \frac{1}{2}(\delta - \lambda_2) \leq i(G) \leq \sqrt{\Delta^2 - \lambda_1^2},$$

where δ and Δ are the minimum and maximum degrees of the graph, respectively, and λ_j are the eigenvalues of the adjacency matrix in a nonincreasing order as before. Consequently, for graphs with bounded maximum and minimum degree—such as the Sierpinski graphs, where $\delta = 2$ and $\Delta = 4$ —the isoperimetric number is determined by the spectral gap $\lambda_1 - \lambda_2$. A large spectral gap indicates a large isoperimetric number, while a small spectral gap indicates a small isoperimetric number. We illustrate in Figure 11(b) the plot of the reciprocal of the density of the Sierpinski graphs against the spectral gap of their adjacency matrices. As can be seen, the Sierpinski graphs with small density, i.e., those with large number of nodes, have very small spectral gaps, and consequently small isoperimetric numbers.

On the contrary, if a graph has a large spectral gap, i.e., $(\lambda_1 - \lambda_2) \rightarrow \infty$, the communicability function is given by

$$(5.10) \quad G_{pq} \rightarrow \psi_{1,p}\psi_{1,q} \exp(\lambda_1),$$

which implies that

$$(5.11) \quad \theta_{pq} \rightarrow 0^\circ \quad \forall p, q \in V.$$

That is, a large isoperimetric number indicates that the graphs have a large spectral gap. At the same time, a large spectral gap indicates that the communicability angle is very small for every pair of nodes in the graph. As we have seen a small spectral gap, and consequently a small isoperimetric number, gives rise to a large communicability angle as in the case of large Sierpinski graphs. This conclusion again supports our idea that the communicability angle is a good indicator of the spatial efficiency of a given network.

5.5. Conclusions of the Computational Analysis of Simple Graphs. The main conclusion of section 5 is that the average communicability angle describes very well a graph characteristic which represents their spatial efficiency. This is drawn from the following observations. First, planar graphs are not spatially efficient graphs, they have large average communicability angles. On the contrary, highly nonplanar graphs more efficiently use the available space, but they have smaller values of $\langle \theta \rangle$. Second, a high modularity graph uses the available space less effectively than a low modularity one, but high modularity graphs have relatively large values of the average communicability angle. Third, graphs containing structural holes, which are not spatially efficient, display large communicability angles, while those having large isoperimetric numbers and consequently good spatial efficiency have communicability angles close to zero.

We should, however, be careful in analyzing more complex situations in which combinations of properties, such as nonplanarity and modularity, or nonplanarity and structural holes, are present. In general, we consider that graphs with relatively small values of the average communicability angle exhibit higher spatial efficiency than those with relatively larger values.

6. Communicability Angle in Real-World Networks. We start this section by considering the average communicability angle of a series of 120 networks arising from

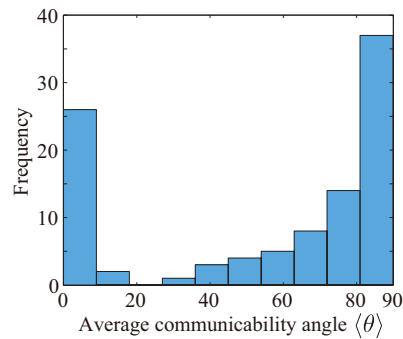


Fig. 12 Histograms of the average communicability angle in 120 real-world networks with a bin size of 9° .

various scenarios. The networks are briefly described in the supplementary information accompanying this paper, where references to the original datasets are provided. The series includes networks in which the nodes and links are clearly embedded into geometrical spaces, such as urban street networks, networks formed by animal nests, brain and neural networks, and protein-residue networks as well as electronic circuits and the Internet. It also includes networks in which the nodes and links can barely be allocated to geographic positions, such as food webs, social networks, and software networks. The biomolecular networks including protein-protein interaction and gene transcription networks are also nongeographically embedded.

6.1. Global Properties of the Communicability Angle. The 120 real-world networks studied here cover the whole spectrum of values of the average communicability angle from $\langle \theta \rangle \approx 10^{-5^\circ}$ for the food web of Shelf to $\langle \theta \rangle \approx 89.9^\circ$ for the Power Grid network of the western United States.

The average communicability angle of these real-world networks is not correlated to the average path length, the communication efficiency, or the resistance distance (see the supplementary information accompanying this paper). To mention an example, let us consider the network of galleries created by ants and the collaboration network associated with the Linux open-source software system (see the supplementary information for details). The first network is planar due to the fact that ants are obliged to create their corridors and galleries in a very thin layer of sand. The second is a highly nonplanar network. Both networks have the communication efficiency $E \approx 0.24$; according to this index the two graphs are equally efficient in transmitting information, something hard to believe taking into account their different topologies and functionalities. The average communicability angle, on the other hand, clearly indicates the fact that the software network has a high spatial efficiency, as is indicated by $\langle \theta \rangle \approx 3.47^\circ$, while the ant network has a very low efficiency, as is indicated by $\langle \theta \rangle \approx 85.51^\circ$.

The histogram in Figure 12 shows two prominent peaks at $0^\circ \leq \langle \theta \rangle \leq 9^\circ$ and $81^\circ \leq \langle \theta \rangle \leq 90^\circ$. A more detailed view (not shown) indicates that the highest frequency occurs at $0^\circ \leq \langle \theta \rangle \leq 1^\circ$, followed by one at $89^\circ \leq \langle \theta \rangle \leq 90^\circ$. That is, the real-world networks are very polarized into the two extremes; either they have very small values of the communicability angle, or very large ones.

Certain classes of networks have a large homogeneity in the values of the average communicability angle. The 1997 and 1998 versions of the Internet at Autonomous

System (AS) have average communicability angles of 0.78° and 0.42° , respectively. There is also a large homogeneity among the brain/neural networks, namely, the visual-cortex networks of cat and macaque as well as the neural network of *C. elegans*, which have $\langle\langle\theta\rangle\rangle = 1.77^\circ \pm 1.66^\circ$, where the double brackets $\langle\langle\cdot\rangle\rangle$ denote the average value of the average communicability angles for a series of networks. In addition, the classes of urban street networks formed by 14 networks and of protein-residue networks formed by 40 networks also show remarkable homogeneity. For instance, the urban street networks have $\langle\langle\theta\rangle\rangle = 86.07^\circ \pm 5.07^\circ$ and the protein-residue networks have $\langle\langle\theta\rangle\rangle = 78.83^\circ \pm 7.28^\circ$. The ranking of the 14 cities in the former is Barcelona < Rio Grande < Yuliang < Chegkan < Atlanta < Berlin < Rotterdam < Hong Kong < Mecca < Cambridge < Oxford < Ahmedabad < Milton Keynes. This means that in terms of the effective communication among the different regions of the city, Barcelona is the most effective, while Milton Keynes is the worst.

The homogeneity among the protein-residue networks is more unexpected than that among the urban street networks because they represent three-dimensional (3D) objects. Proteins are folded into 3D structures forming topologies consisting of a mix of α -helices and β -sheets. They also have different shapes and sphericities. It is therefore surprising that the protein-residue networks are characterized by very large values of the communicability angle, which are more characteristic of planar or almost planar networks, as demonstrated for the urban street networks.

Although we will return below to the relationship between the communicability angle and the structure of proteins, let us make a comment here. The fact that proteins are embedded into 3D physical space does not necessarily mean that their residue networks are nonplanar. The same applies to other naturally evolving networks, such as the networks of galleries and corridors formed by termite mounds, which are also characterized by very large average communicability angles with $\langle\langle\theta\rangle\rangle = 88.33^\circ \pm 1.01^\circ$. Although the mounds are constructed in 3D space, they are remarkably close to planar graphs; we have indeed found that by removing only 6% of the edges of these networks the graphs representing them become planar. Both the termite mounds and the protein-residue networks have certainly evolved in 3D space, but the networks must be close to planar graphs for different ecological or biological reasons. In the termite mounds a large volume of the 3D space must be used to produce the ventilation system necessary to discharge the carbon dioxide produced in its interior. For proteins, structures close to planar ones are needed to avoid the high compactness that destroys the internal cavities of the proteins needed for developing their functions; see section 6.2 below.

On the other hand, the values of $\langle\theta\rangle$ obtained for the software networks [31] are unexpectedly heterogeneous. These networks yield $\langle\langle\theta\rangle\rangle = 57.6^\circ \pm 30.7^\circ$ with values ranging from $\langle\theta\rangle \approx 3.465^\circ$ for Linux to $\langle\theta\rangle \approx 84.323^\circ$ for XMMS. The ranking of these networks in terms of the average communicability angle is Linux < MySQL < VTK < Abi Word < Digital Material < XMMS. The classes of social and biological networks consisting of 14 and 11 networks, respectively, also show relatively large variability in their values of the communicability angle: $\langle\langle\theta\rangle\rangle = 55.8^\circ \pm 21.3^\circ$ and $\langle\langle\theta\rangle\rangle = 63.3^\circ \pm 17.0^\circ$, respectively. This is not surprising; we can easily associate it with the diversity of networks in these classes.

What is really surprising is that the food webs, which form a very homogeneous class of networks in terms of the relationships accounted for them, yield a relatively large standard deviation in the values of the communicability angle: $\langle\langle\theta\rangle\rangle = 7.1^\circ \pm 16.1^\circ$ with values ranging from $\langle\theta\rangle \approx 10^{-5^\circ}$ for the marine system of Shelf to $\langle\theta\rangle \approx 78.356^\circ$ for the web of the English grassland. The ranking of these food webs in terms of

the average communicability angle is Shelf < Elverde < Skipwith < ReefSmall < LittleRock < Stony < Coachella < Canton < Benguela < BridgeBrook < Ythan2 < Ythan1 < StMartins < StMarks < ScotchBroom < Chesapeake < Grassland.

In terms of the individual values of $\langle \theta \rangle$, the results obtained for these 120 networks agree with our findings in the previous section. The largest average communicability angles are observed for the Power Grid of the western United States and urban street networks, which are planar or almost planar with both nodes and edges embedded into a plane. On the other extreme of the smallest average communicability angles, there are networks which are highly nonplanar, such as the U.S. air transportation network, a world trade network, the Internet at AS, and brain/neural networks. All these networks have nodes embedded into 2D or 3D spaces, such as cities, countries, or organs, but the edges connecting them very efficiently use the available space. We remark here that the small values of $\langle \theta \rangle$ observed in some classes of networks do not necessarily mean a high interconnection density. For instance, the U.S. airport transportation network and the two versions of the Internet studied here have relatively small edge densities, 0.039 and 0.0011, respectively.

6.2. Communicability Angle and Spatial Efficiency of Proteins. We have accumulated several pieces of empirical evidence that support the idea that the average communicability angle accounts for the spatial efficiency of graphs. It is, however, generally difficult to find quantitative measures of the spatial efficiency in real-world networks to compare with the communicability angle.

An exception to this is provided by proteins, which are 3D objects characterized by different degrees of packing or spatial efficiency. In this section we study the relationship between the average communicability angle and the spatial efficiency of the protein-residue networks for a group of 40 proteins whose 3D structures have been resolved by X-ray crystallography and deposited in the protein databank (PDB) [5]. Here each node represents an amino acid in the protein and two nodes are connected if the corresponding amino acids are separated by a distance of no more than 7Å in the 3D structure of the protein, as determined experimentally [3].

A protein is a linear sequence of amino acids connected by peptide bonds. The chain is folded into a 3D shape unique to each protein. While the amino-acid sequence forms the so-called primary structure of the protein, the 3D folding defines its secondary and tertiary structures. The secondary structure is characterized by the presence of α -helices and β -sheets, while the tertiary structure is formed by global positioning of the secondary structure into a 3D shape that gives the protein its globular-like structure [10]. The folding of the proteins is the consequence, *grosso modo*, of two main necessities that the protein has: (i) protecting the hydrophobic amino acids from their contact with water; and (ii) occupying a minimum space inside the limited volume of the cell. Thus, the packing of a protein is related to its spatial efficiency [20], which is responsible for many of its physico-chemical and biological properties.

There are many ways of quantifying the packing of a protein, but here we consider the following one. Let V_e be the volume of a protein which is expected from its ideal 3D structure and let V_o be the volume which is actually observed in its X-ray crystallography. We then define the relative deviation from its ideal volume as

$$(6.1) \quad P = \frac{V_e - V_o}{V_e}.$$

Hereafter, we call P the relative packing efficiency of the protein. A positive value of

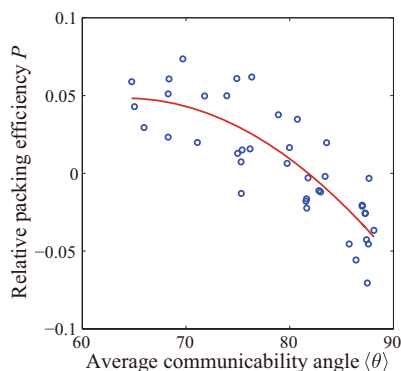


Fig. 13 Linear correlation between the average communicability angle of proteins represented by residue networks and the relative packing efficiency.

P means that the protein is more packed than is expected from its ideal 3D structure, that it is highly efficient in using the 3D space, at least relative to the ideal structure. A negative value of P , on the other hand, means that it is less packed than expected, that it is not spatially efficient. We should mention here that values that deviate strongly from the expected or ideal values can indicate possible problems with the structure and as such should be discarded from the analysis.

Using computational techniques and VADAR software described in [38], we have calculated the expected and observed volumes of the 40 proteins. We show in Figure 13 the relationship between the relative packing efficiency P and the average communicability angle of the 40 proteins. The Pearson correlation coefficient is $R = -0.837$, indicating a significant correlation between the two variables. We can summarize the results as follows: (i) proteins with poor spatial efficiency, $P < 0$, have $\langle \theta \rangle > 81^\circ$; (ii) those with high spatial efficiency, $P > 0$, have $\langle \theta \rangle < 80^\circ$. In other words, small average communicability angles are related to high spatial efficiency of proteins, while large average communicability angles indicate a poor use of space. We note in passing that there are no proteins with $\langle \theta \rangle < 60^\circ$, which can be explained by the fact that too much packing would make the internal cavities of the protein disappear [20]. The internal cavities are responsible for the interaction of proteins with other biological molecules and usually play a fundamental role in their functionality. In general, we can conclude that proteins are spongy in a similar way to the Sierpinski graphs.

Possibilities which the communicability angle brings to the analysis of the structure of spatially embedded networks obviously go beyond the use of $\langle \theta \rangle$. For instance, the contour plot of the communicability angle for every pair of residues in a protein can reveal important properties of its 3D structure. Figure 14 shows an example of the protein with PDB code 1amm, which corresponds to the GammaB crystalline, whose crystallographic analysis was carried out at 150K. This protein consists of two α , β -domains, the first of which is formed by amino acids 1–83 and the second by amino acids 84–174. The two domains are very well reflected in the contour plot Figure 14(a) as two main diagonal blocks of relatively small communicability angles, which indicates good internal communication in each domain.

6.3. Spatial Efficiency in Networks under External Stress. The communicability function has been previously generalized to consider an external stress to which the

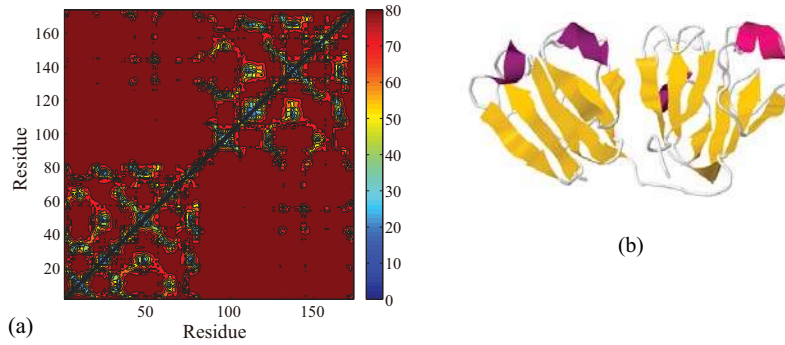


Fig. 14 (a) Contour plot of the communicability angle between every pair of residues in the GammaB crystalline protein with PDB code 1amm. (b) A cartoon representation of the protein with PDB code 1amm in which the β -sheets are represented as arrows in yellow and the helices as ribbons in magenta. This shows the existence of two domains.

network is submitted. This external stress is accounted for by means of the so-called inverse temperature $\beta \equiv (k_B T)^{-1}$, where k_B is a constant and T is the temperature [14]. This analogy results from supposing that the whole network is submerged in a thermal bath of the inverse temperature β ; see [16, 11] for details. After equilibration in the bath, all edges of the network acquire a weight equal to β . In other words, we use the weighted matrix βA , where all edges have the weight β , instead of the adjacency matrix, where the edge weights are ignored.

It is clear that when $\beta \rightarrow 0$, i.e., as the temperature tends to infinity, the network becomes disconnected and there is no communication between any pair of nodes. This resembles a gas in which every node is an independent particle. On the other hand, when $\beta \rightarrow \infty$, i.e., the temperature tends to zero, the weights of every edge become extremely large, which definitively increases the communication capacity among the pairs of connected nodes. The temperature thus plays the role of an empirical parameter which is useful in simulating effects of external stresses to which the network is submitted, such as different levels of social agitation, economical situations, environmental stress, variable physiological conditions, etc. Under this analogy, we generalized the communicability function (2.2) into the form [14]

$$(6.2) \quad G_{pq}(\beta) = (e^{\beta A})_{pq}.$$

It is straightforward to realize that the communicability angle between a given pair of nodes is generalized to

$$(6.3) \quad \cos \theta_{pq}(\beta) = \frac{G_{pq}(\beta)}{\sqrt{G_{pp}(\beta) G_{qq}(\beta)}}.$$

Let us conduct a simple experiment to explore the possibilities which this empirical parameter brings to the analysis of real-world scenarios. We use two urban street networks representing the city landscapes of Rio Grande in Brazil and Yuliang in China. Both cities have large values of the average communicability angle, i.e., low spatial efficiency, with $\langle \theta \rangle \approx 79.7^\circ$ and $\langle \theta \rangle \approx 85.8^\circ$, respectively. We then lower the temperature and see if it increases the spatial efficiency of both cities, i.e., if it decreases the values of $\langle \theta \rangle$. In other words, we systematically increase β and compute the average communicability angle $\langle \theta(\beta) \rangle$. The increase in β here can be associated with the average increment in the number of lanes per street in the city.

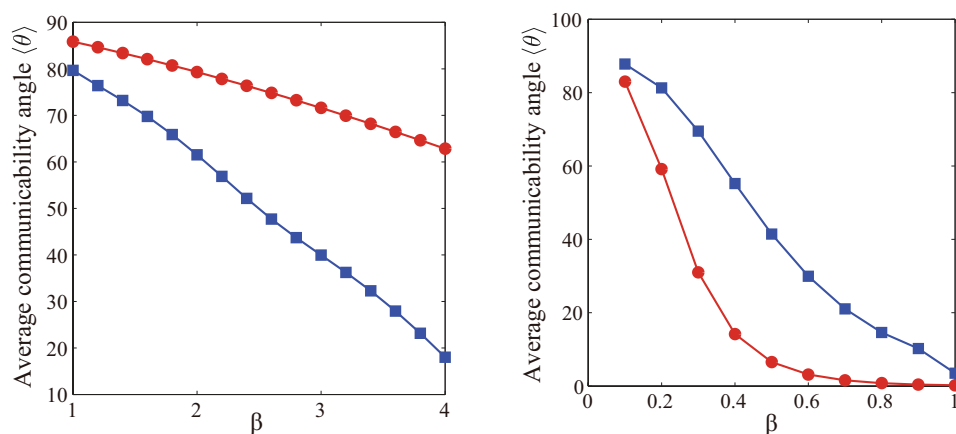


Fig. 15 (a) Effects of the inverse temperature β on the average communicability angle in two urban street networks, Rio Grande, Brazil (squares) and Yuliang, China (circles). (b) The same for two visual-cortex networks: cat (circles) and macaque (squares).

Figure 15(a) shows the results. The city of Rio Grande dramatically improves its spatial efficiency by increasing the average number of lanes of its streets. Although the improvement for Yuliang is not so dramatic, there is still a decrease in the average communicability angle of 20° . The causes for the difference in the variation of $\langle \theta \rangle$ with the temperature for different networks are not trivial, as there are likely to be many structural factors involved. We do not investigate those causes here.

Here we should point out a metaphor that can help us to understand how the change in the parameter β can change the spatial efficiency of a network. Suppose that we analyze an urban street network formed only by streets of just one lane each. The spatial efficiency of that urban street network can be dramatically improved if we now change all the streets to have $k > 1$ lanes. This is exactly what the parameter β is doing to the network and the result is the intuitive increase in spatial efficiency of the whole system.

We next carry out the opposite experiment using two brain networks representing the cat and macaque visual cortices. The average communicability angle shows that both networks have a great spatial efficiency, $\langle \theta \rangle \approx 0.22^\circ$ and $\langle \theta \rangle \approx 3.52^\circ$, respectively. We now raise the temperature, i.e., decrease β , and see if it deteriorates the connections in the visual cortices in terms of the average communicability angle $\langle \theta(\beta) \rangle$. The decrease of β can be regarded as malfunctioning or disease.

Figure 15(b) shows the results. Both networks dramatically decrease their spatial efficiency as $\beta \rightarrow 0$; obviously, $\theta_{pq}(\beta = 0) = 90^\circ$. We notice, however, that the cat visual cortex is more resistant to the stress than that of the macaque. For $\beta = 0.6$, for example, the former has $\langle \theta \rangle \approx 3.15^\circ$, while the latter has jumped up to $\langle \theta \rangle \approx 29.9^\circ$.

The influence of the inverse temperature can be summarized as follows. In the limit $\beta \rightarrow \infty$, we have $G_{pq} \rightarrow \psi_{1,p}\psi_{1,q} \exp(\beta\lambda_1)$. This corresponds to the increase of the good expansion properties of the network. We recall from previous sections that for expanders the spectral gap $\lambda_1 - \lambda_2$ is very large, and consequently we have the convergence above. This is exactly the effect that we see; when we increase β , the networks become more spatially efficient, i.e., $\langle \theta \rangle \rightarrow 0^\circ$. In the limit $\beta \rightarrow 0$, on the other hand, we have $G_{pq} \rightarrow 1$, which implies that $\langle \theta \rangle \rightarrow 90^\circ$. This is equivalent to

dramatically reducing the capacity of each edge of transmitting information in the network, which clearly decreases its communication and spatial efficiencies.

In closing, the use of the empirical parameter β allows us to simulate the effects of external factors which can modify the spatial efficiency of a network. This brings a modeling scenario to the assaying of strategies to improve the spatial efficiency of networks or to analyses of their resilience to external stresses.

7. Conclusions. In the present paper, we introduced a new measure called the communicability angle for networks. We have showed that this captures the spatial efficiency of networks. In a network, the more abstract spatial efficiency refers to the average quality of communication among the nodes. Such communication goodness is quantified as the ratio of the amount of information successfully delivered to its destination to the information which is frustrated in its delivery and returned to its originators. This new paradigm is then mathematically formulated in terms of the communicability angle between a pair of nodes. We have provided analytical and empirical pieces of evidence which reaffirm the idea that the communicability angle accounts for the spatial efficiency of networks.

The richness of this approach goes beyond the results presented here; there are a few immediate directions of research in this area which can open new opportunities for the analysis of networks. The use of the communicability angle for a pair of connected nodes can be seen as an edge centrality measure which may reveal important characteristics of individual edges in networks. The communicability angle averaged over the edges incident to a given node can also represent a node centrality index which indicates the contribution of the node to the global spatial efficiency of a network. The study of the effects of the inverse temperature on the spatial efficiency and the determination of the most important structural factors that influence it is of tremendous practical importance. These studies will allow us not only to predict the effects of external stresses over the spatial efficiency of a network, but also to assay theoretical scenarios of improving this efficiency in certain classes of networks. Last but not least, the new concept of communicability angle can bring new possibilities to the mathematical analysis of specific types of graphs and properties, such as planarity and graph thickness, among others.

Acknowledgment. EE thanks the Royal Society for a Wolfson Research Merit Award. He also thanks Dr. Sean Hanna (UCL) for the datasets of urban street networks used in this work.

REFERENCES

- [1] S. ACHARD AND E. BULLMORE, *Efficiency and cost of economical brain functional networks*, PLoS Comput. Biol., 3 (2007), e17. (Cited on p. 693)
- [2] N. ALON AND V. D. MILMAN, λ_1 , *isoperimetric inequalities for graphs, and superconcentrators*, J. Combin. Theory Ser. B, 38 (1985), pp. 73–88. (Cited on p. 702)
- [3] A. R. ATILGAN, P. AKAN, AND C. BAYSAL, *Small-world communication of residues and significance for protein dynamics*, Biophys. J., 86 (2004), pp. 85–91. (Cited on p. 709)
- [4] M. BARTHÉLEMY, *Spatial networks*, Phys. Rep., 499 (2011), pp. 1–101. (Cited on p. 692)
- [5] H. M. BERMAN, J. WESTBROOK, Z. FENG, G. GILLILAND, T. N. BHAT, H. WEISSIG, I. N. SHINDYALOV, AND P. E. BOURNE, *The protein data bank*, Nucleic Acids Res., 28 (2000), pp. 235–242. (Cited on p. 709)
- [6] J. M. BOYER AND W. J. MYRVOLD, *On the cutting edge: Simplified $O(n)$ planarity by edge addition*, J. Graph Algorith. Appl., 8 (2004), pp. 241–273. (Cited on p. 700)
- [7] E. BULLMORE AND O. SPORNS, *Complex brain networks: Graph theoretical analysis of structural and functional systems*, Nature Rev. Neurosci., 10 (2009), pp. 186–198. (Cited on p. 692)

- [8] A. CARDILLO, S. SCELLATO, V. LATORA, AND S. PORTA, *Structural properties of planar graphs of urban street patterns*, Phys. Rev. E, 73 (2006), 066107. (Cited on p. 692)
- [9] L. F. COSTA, O. N. OLIVEIRA, JR., G. TRAVIESO, F. A. RODRIGUES, P. R. VILLAS BOAS, L. ANTIQUEIRA, M. P. VIANA, AND L. E. CORREA ROCHA, *Analyzing and modeling real-world phenomena with complex networks: A survey of applications*, Adv. Phys., 60 (2011), pp. 329–412. (Cited on p. 692)
- [10] E. ESTRADA, *Characterization of the folding degree of proteins*, Bioinformatics, 18 (2002), pp. 697–704. (Cited on pp. 692, 709)
- [11] E. ESTRADA, *The Structure of Complex Networks. Theory and Applications*, Oxford University Press, Oxford, UK, 2011. (Cited on pp. 692, 693, 711)
- [12] E. ESTRADA, *The communicability distance in graphs*, Linear Algebra Appl., 436 (2012), pp. 4317–4328. (Cited on pp. 694, 695)
- [13] E. ESTRADA, *Complex networks in the Euclidean space of communicability distances*, Phys. Rev. E, 85 (2012), 066122. (Cited on p. 694)
- [14] E. ESTRADA AND N. HATANO, *Statistical-mechanical approach to subgraph centrality in complex networks*, Chem. Phys. Lett., 439 (2007), pp. 247–251. (Cited on p. 711)
- [15] E. ESTRADA AND N. HATANO, *Communicability in complex networks*, Phys. Rev. E, 77 (2008), 036111. (Cited on pp. 693, 694)
- [16] E. ESTRADA, N. HATANO, AND M. BENZI, *The physics of communicability in complex networks*, Phys. Rep., 514 (2012), pp. 89–119. (Cited on pp. 693, 694, 711)
- [17] E. ESTRADA AND D. J. HIGHAM, *Network properties revealed through matrix functions*, SIAM Rev., 52 (2010), pp. 696–714, doi:10.1137/090761070. (Cited on pp. 693, 694)
- [18] E. ESTRADA AND J. A. RODRÍGUEZ-VELÁZQUEZ, *Subgraph centrality in complex networks*, Phys. Rev. E, 71 (2005), 056103. (Cited on p. 694)
- [19] E. ESTRADA, M. G. SANCHEZ-LIROLA, AND J. A. DE LA PEÑA, *Hyperspherical embedding of graphs and networks in communicability spaces*, Discrete Appl. Math., 176 (2014), pp. 53–77. (Cited on p. 695)
- [20] P. J. FLEMING AND F. M. RICHARDS, *Protein packing: Dependence on protein size, secondary structure and amino acid composition*, J. Mol. Biol., 299 (2000), pp. 487–498. (Cited on pp. 709, 710)
- [21] J. GOŃI, A. AVENA-KOENIGSBERGER, N. V. DE MENDIZABAL, M. VAN DEN HEUVEL, R. BETZEL, AND O. SPORNS, *Exploring the morphospace of communication efficiency in complex networks*, PLoS One, 8 (2013), e58070. (Cited on p. 693)
- [22] R. L. GRAHAM AND N. J. A. SLOANE, *Penny packing and two-dimensional codes*, Discrete Comput. Geom., 5 (1990), pp. 1–11. (Cited on p. 704)
- [23] J. L. GROSS AND T. W. TUCKER, *Topological Graph Theory*, Dover, Mineola, NY, 1987. (Cited on p. 693)
- [24] S. HOORY, N. LINIAL, AND A. WIGDERSON, *Expander graphs and their applications*, Bull. Amer. Math. Soc., 43 (2006), pp. 439–561. (Cited on p. 702)
- [25] B. JIANG AND C. CLARAMUNT, *Topological analysis of urban street networks*, Environ. Plan. B, 31 (2004), pp. 151–162. (Cited on p. 692)
- [26] D. J. KLEIN AND M. RANDIĆ, *Resistance distance*, J. Math. Chem., 12 (1993), pp. 81–95. (Cited on p. 694)
- [27] V. LATORA AND M. MARCHIORI, *Efficient behavior of small-world networks*, Phys. Rev. Lett., 87 (2001), 198701. (Cited on pp. 693, 694)
- [28] V. LATORA AND M. MARCHIORI, *Economic small-world behavior in weighted networks*, Eur. Phys. J. B, 32 (2003), pp. 249–263. (Cited on p. 693)
- [29] A. LUBOTZKY, *Expander graphs in pure and applied mathematics*, Bull. Amer. Math. Soc., 49 (2012), pp. 113–162. (Cited on p. 702)
- [30] B. MOHAR, *Isoperimetric inequalities, growth, and spectrum of graphs*, Linear Algebra Appl., 103 (1983), pp. 119–131. (Cited on p. 705)
- [31] C. R. MYERS, *Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs*, Phys. Rev. E, 68 (2003), 046116. (Cited on p. 708)
- [32] M. E. J. NEWMAN, *The structure and function of complex networks*, SIAM Rev., 45 (2003), pp. 167–256, doi:10.1137/S003614450342480. (Cited on p. 692)
- [33] M. E. J. NEWMAN, *Modularity and community structure in networks*, Proc. Natl. Acad. Sci. USA, 103 (2006), pp. 8577–8582. (Cited on p. 702)
- [34] P. SARNAK, *What is... an expander?*, Notices Amer. Math. Soc., 51 (2004), pp. 762–763. (Cited on p. 702)
- [35] A. SARZYŃSKI AND A. LEVY, *Spatial Efficiency and Regional Prosperity: A Literature Review and Policy Discussion*, prepared as background for GWIPP’s Implementing Regionalism project, funded by the Surdna Foundation, <https://gwipp.gwu.edu/files/downloads/>

- Working_Paper_041_RegionalProsperity.pdf (accessed 20 November 2014). (Cited on p. 693)
- [36] W. SCHNYDER, *Embedding planar graphs on the grid*, in Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '90), SIAM, Philadelphia, 1990, pp. 138–148. (Cited on p. 701)
 - [37] E. TEUFL AND S. WAGNER, *The number of spanning trees of finite Sierpinski graphs*, in Proceedings of the Fourth Colloquium on Mathematics and Computer Science, DMTCS Proc. AG, 2006, pp. 411–414. (Cited on p. 704)
 - [38] L. WILLARD, A. RANJAN, H. ZHANG, H. MONZAVI, R. F. BOYKO, B. D. SYKES, AND D. S. WISHART, *VADAR: A web server for quantitative evaluation of protein structure quality*, Nucleic Acids Res., 31 (2003), pp. 3316–3319. (Cited on p. 710)