Linear Algebra and its Applications 436 (2012) 4317–4328

Contents lists available at SciVerse ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

The communicability distance in graphs

Ernesto Estrada

Department of Mathematics and Statistics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XQ, UK

ARTICLE INFO

Article history:
Received 12 January 2012
Accepted 24 January 2012
Available online 25 February 2012

Submitted by R. Brualdi

AMS classification:
05C12
05C50
15A18
15A16

Keywords:
Euclidean distance
Graph spectrum
Graph distance
Communicability
Matrix functions

ABSTRACT

Let G be a simple connected graph with adjacency matrix $\mathbf{A}$. The communicability G_{pq} between two nodes p and q of the graph is defined as the pq-entry of $\mathbf{G} = \exp(\mathbf{A})$. We prove here that $\xi_{p,q} = (G_{pp} + G_{qq} - 2G_{pq})^{1/2}$ is a Euclidean distance and give expressions for it in paths, cycles, stars and complete graphs with n nodes. The sum of all communicability distances in a graph is introduced as a new graph invariant $\Upsilon(G)$. We compare this index with the Wiener and Kirchhoff indices of graphs and conjecture about the graphs with maximum and minimum values of this index.

1. Introduction

Through this paper we only consider simple connected graphs, i.e., finite, undirected connected graphs without loops and multiple edges. Let G be one of such graphs with node set $V(G) = \{v_1, v_2, \ldots, v_n\}$. The communicability between the nodes p and q in G has been defined as a weighted sum of all walks starting at node p and ending at node q, in which the weighting scheme gives more weight to the shortest walks than to the longer ones [1]. A walk of length k is a sequence of (not necessarily distinct) nodes $v_0, v_1, \ldots, v_{k-1}, v_k$ such that for each $i = 1, 2, \ldots, k$ there is an edge from v_{i-1} to v_i. Let $\mathbf{A}$ be the adjacency matrix of a simple graph on n nodes. Then, the communicability is defined as follow. Let

$$\mathbf{G} \overset{def}{=} \sum_{k=0}^{\infty} \frac{\left(\mathbf{A}^k\right)}{k!} = e^{\mathbf{A}}. \tag{1}$$

E-mail address: ernesto.estrada@strath.ac.uk

0024-3795/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.laa.2012.01.017

Then, the communicability between the nodes p and q is defined as [1] the corresponding nondiagonal entry of $\mathbf{G}$, G_{pq}. If $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ are the eigenvalues of $\mathbf{A}$ we have

$$G_{pq} = \sum_{j=1}^{n} \varphi_j(p)\varphi_j(q)e^{\lambda_j}, \tag{2}$$

where $\varphi_j(p)$ and $\varphi_j(q)$ are the pth and qth entries of the jth orthonormal eigenvector of $\mathbf{A}$ associated with the eigenvalue λ_j. The diagonal entries of $\mathbf{G}$ are the so-called *self-communicability* or the '*subgraph centrality*' of the corresponding node [2]. The term 'subgraph centrality' refers to the fact that G_{pp} counts the weighted participation of the node p in all subgraphs of a graph [2]. The self-communicability was first proposed in 2000 for the study of the degree of folding of proteins [3]. Since then, the expression (1) has been reinterpreted as follows [1,4,5]. Let us consider a network as a balls-and-springs system in which every node is represented by a ball of mass m and every edge is a spring with the spring constant $m\omega^2$, where ω is the angular frequency, connecting two balls. The network is considered to be submerged into a thermal bath at the temperature T. Then the balls in the graph oscillate under thermal disturbances. The Hamiltonian of the oscillator network has the form

$$H = \sum_i \left[\frac{p_i^2}{2m} + (K - k_i)\,\frac{m\omega^2 x_i^2}{2} \right] + \frac{m\omega^2}{2} \sum_{\substack{i,j \\ (i<j)}} A_{ij}\,(x_i - x_j)^2, \tag{3}$$

where k_i is the degree of the node i, K is a constant satisfying $K \geqslant \max_i k_i$, x_i is the coordinate of the ball i, which indicates the fluctuation of the ball i from its equilibrium point $x_i = 0$. The second term of the right-hand side is the potential energy of the springs connecting the balls, because $x_i - x_j$ is the extension or the contraction of the spring connecting the nodes i and j. The first term in the first square parentheses is the kinetic energy of the ball i, whereas the second term in the first square parentheses is a counter term that offsets the movement of the network as a whole by tying the network to the ground. We add this term because we are only interested in small oscillations around the equilibrium [5]. We consider that the network obeys the laws of quantum mechanics, basically that the momenta p_j and the coordinates x_i are not independent variables but they are operators that satisfy the commutation relation, $[x_i, p_j] = i\hbar\delta_{ij}$, where δ_{ij} is the Dirac function, $i = \sqrt{-1}$ and $\hbar$ is the Dirac constant. In this case, it is proved analytically that G_{pq} is the thermal Green's function of the network of quantum harmonic oscillators when the inverse temperature of the system is equal to one. Consequently, G_{pp} indicates how much of an excitation at the node p propagates through the network before coming back to the same node and being annhilated. G_{pq} measures how much of such excitation is transmitted from p to q. Both measures have found many applications in the study of complex systems and the reader is referred to the recent review [5] for details and references.

In this paper we are going to define a Euclidean distance between the nodes of a graph based on the concept of communicability. The best known distance in graphs is the shortest path or geodesic distance [6]. The shortest path distance d_{pq} between the nodes p and q is defined as the number of edges in the shortest path connecting both nodes. A path connecting node p to node q is a pq-walk in which all nodes and edges are different. Another graph distance is the so-called resistance or commute distance [7–9]. If $\mathbf{L}^+$ is the generalized Moore–Penrose inverse of the Laplacian matrix $\mathbf{L} = \mathbf{K} - \mathbf{A}$, where $\mathbf{K}$ is the diagonal matrix of node degrees, the resistance distance between nodes p and q is given by $\Omega_{pq} = (\mathbf{L}^+)_{pp} + (\mathbf{L}^+)_{qq} - 2\,(\mathbf{L}^+)_{pq}$. More recently, Chebotarev has defined distances that generalize the shortest path and resistance distances in graphs [10–12]. The reader is referred to the original literature for details about these graph distances.

2. Communicability distance

We start by considering that there is a simultaneous excitation at two nodes p and q of a graph. We know that G_{pp} and G_{qq} measure the amount of excitation which returns to the respective nodes and

G_{pq} quantifies the amount of such excitation transmitted from one node to another. Then,

$$\xi_{pq}^2 \stackrel{def}{=} G_{pp} + G_{qq} - 2G_{pq} \tag{4}$$

accounts for the differences on the amount of excitation that returns to the nodes to the one transmitted between them. The reason for the use of the square will be clear immediately when we prove the main result of this paper.

Theorem 1. *ξ_{pq} is a Euclidean distance between the nodes p and q of the graph.*

Proof. We start by recalling that a Euclidean metric d_E is the metric on $\mathbb{R}^n$ defined by $d_E = \|x - y\|_2 = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$ (see [13, p. 94]). Now, let $\mathbf{\Lambda}$ be a diagonal matrix of the eigenvalues of the adjacency matrix and let

$$\boldsymbol{\varphi}_p = \left[\varphi_1(p) \ \varphi_2(p) \ \cdots \ \varphi_n(p) \right]^T, \tag{5}$$

which can be obtained by transposing the pth row of the matrix $\mathbf{U}$ of eigenvectors of the adjacency matrix. Then, we can write (4) as

$$\xi_{pq}^2 = \left(\boldsymbol{\varphi}_p - \boldsymbol{\varphi}_q\right)^T e^{\mathbf{\Lambda}} \left(\boldsymbol{\varphi}_p - \boldsymbol{\varphi}_q\right), \tag{6}$$

which can be regrouped as

$$\begin{aligned} \xi_{pq}^2 &= \left[e^{\mathbf{\Lambda}/2}\left(\boldsymbol{\varphi}_p - \boldsymbol{\varphi}_q\right)\right]^T e^{\mathbf{\Lambda}/2}\left(\boldsymbol{\varphi}_p - \boldsymbol{\varphi}_q\right) \\ &= \left(e^{\mathbf{\Lambda}/2}\boldsymbol{\varphi}_p - e^{\mathbf{\Lambda}/2}\boldsymbol{\varphi}_q\right)^T \left(e^{\mathbf{\Lambda}/2}\boldsymbol{\varphi}_p - e^{\mathbf{\Lambda}/2}\boldsymbol{\varphi}_q\right). \end{aligned} \tag{7}$$

Now, we can define the vector $\mathbf{x}_p = e^{\mathbf{\Lambda}/2}\boldsymbol{\varphi}_p$ and show that ξ_{pq}^2 is a square norm

$$\begin{aligned} \xi_{pq}^2 &= \left(\mathbf{x}_p - \mathbf{x}_q\right)^T \left(\mathbf{x}_p - \mathbf{x}_q\right) \\ &= \left\|\mathbf{x}_p - \mathbf{x}_q\right\|^2, \end{aligned} \tag{8}$$

which obviously means that

$$\xi_{pq} = \sqrt{\left\|\mathbf{x}_p - \mathbf{x}_q\right\|^2} = \sqrt{\left\|\mathbf{x}_p\right\|^2 + \left\|\mathbf{x}_q\right\|^2 - 2\mathbf{x}_p \cdot \mathbf{x}_q} \tag{9}$$

is a Euclidean distance between the nodes p and q of the graph. □

We will call ξ_{pq} the communicability distance between the nodes p and q of a graph. We should notice here that similar graph distances can be obtained from any positive definite matrix related to the graph. However, the communicability distance has a clear 'physical' or 'structural' interpretation, which makes it very suitable for the analysis of large complex networks as well as small graphs.

Then, it is known from the theory of Euclidean distances [14–17] that ξ_{pq} has the following properties:

(1) $\xi_{pq} \geqslant 0$ (non-negativity);
(2) $\xi_{pq} = 0 \Leftrightarrow p = q$ (self-distance);
(3) $\xi_{pq} = \xi_{qp}$ (symmetry);
(4) $\xi_{pq} \leqslant \xi_{pr} + \xi_{rq}$ (triangle inequality);
(5) $\cos\left(\theta_{pqr} + \theta_{rqs}\right) \leqslant \cos\theta_{pqs} \leqslant \cos\left(\theta_{pqr} - \theta_{rqs}\right)$, $0 \leqslant \theta_{pqr}, \theta_{rqs}, \theta_{pqs} \leqslant \pi$, where $\theta_{pqr} = \theta_{rqp}$ is the angle formed by the nodes pqr, centered at node q, which is defined as

$$\cos\theta_{pqr} = \frac{\xi_{pq} + \xi_{qr} - \xi_{pr}}{2\sqrt{\xi_{pq}\xi_{qr}}}. \tag{10}$$

Now we show some results about the communicability distance of some elementary graphs, which will help us to interpret this measure when applied to more complex structures. In particular we study the n-nodes path P_n, the n-nodes cycle C_n, the star graph $S_{1,n-1}$, and the complete graph K_n of n nodes. P_n is a connected graph in which $n-2$ nodes are connected to other two nodes and two nodes are connected to only one node; C_n is the connected graph of n nodes in which every node is connected to two others; $S_{1,n-1}$ is the connected graph in which there is one node connected to $n-1$ nodes, here labeled as 1 and named the central node, and $n-1$ nodes are connected to the central one only; and K_n is the graph in which every pair of nodes is connected by an edge. Let $I_\gamma(z)$ be the Bessel function of the first order defined by the following integral [18]:

$$I_\gamma(z) = \frac{1}{\pi}\int_0^\pi \exp(z\cos\phi)\cos(\gamma\phi)\,d\phi - \frac{\sin(\gamma\pi)}{\pi}\int_0^\infty \exp(-z\cosh t - \gamma t)\,dt. \tag{11}$$

Lemma 2. *Let P_n be a path of n nodes labeled as $1, 2, \ldots, n$ staring from any of the two endpoints, and let $p \geqslant q$. Let,*

$$\xi'_{pq} = \sqrt{2I_0(2) - I_{2r(p)}(2) - I_{2r(q)}(2) + 2I_{p+q}(2) - 2I_{p-q}(2)}, \tag{12}$$

where

$$r(i) = \begin{cases} i & \text{if } i \leqslant n/2 (n \text{ even}) \text{ or } i \leqslant (n+1)/2 (n \text{ odd}) \\ n-i+1 & \text{if } i > n/2 (n \text{ even}) \text{ or } i > (n+1)/2\ (n \text{ odd}) \end{cases}.$$

Then, the communicability distance between two nodes of P_n tends to ξ'_{pq} as the size of the path increases, i.e., $\xi_{pq}/\xi'_{pq} \to 1$ as $n \to \infty$.

Proof. The pth entry of the jth eigenvector of the adjacency matrix of P_n is given by $\varphi_j(p) = \sqrt{\frac{2}{n+1}}\sin\frac{jp\pi}{n+1}$ $(j = 1, \ldots, n)$ and the corresponding eigenvalue by $\lambda_j = 2\cos\frac{j\pi}{n+1}$ [19]. Then, the communicability between any two nodes in P_n is given by

$$\begin{aligned} G_{pq} &= \frac{2}{n+1}\sum_{j=1}^{n}\sin\left(\frac{jp\pi}{n+1}\right)\sin\left(\frac{jq\pi}{n+1}\right)e^{2\cos\left(\frac{j\pi}{n+1}\right)} \\ &= \frac{1}{n+1}\sum_{j=1}^{n}\left[\cos\left(\frac{j\pi(p-q)}{n+1}\right) - \cos\left(\frac{j\pi(p+q)}{n+1}\right)\right]e^{2\cos\left(\frac{j\pi}{n+1}\right)}. \end{aligned} \tag{13}$$

Let,

$$\begin{aligned} G'_{pq} &= \frac{1}{\pi}\int_0^\pi \cos[\theta(p-q)]e^{2\cos\theta}d\theta - \frac{1}{\pi}\int_0^\pi \cos[\theta(p+q)]e^{2\cos\theta}d\theta \\ &= I_{p-q}(2) - I_{p+q}(2), \end{aligned} \tag{14}$$

where $\theta = \pi j/(n+1)$ and $I_\gamma(z)$ is the Bessel function of the first order (11). The relationship between G'_{pq} and G_{pq} can be seen as the one existing in numerical methods like the trapezium or Simpson's rules in which an integral is approximated by a summation. Here the analogous of the number of strips used in those numerical methods of integration is the number of nodes in the path. Then, it is easy to realize that $G_{pq}/G'_{pq} \to 1$ as $n \to \infty$.

The expression for the self-communicability of a given node is

$$G_{pp} = \frac{2}{n+1}\sum_{j=1}^{n}\sin^2\left(\frac{j\pi p}{n+1}\right)e^{2\cos\left(\frac{j\pi}{n+1}\right)}$$

$$= \frac{1}{n+1}\sum_{j=1}^{n}\left[1-\cos\left(\frac{2j\pi p}{n+1}\right)\right]e^{2\cos\left(\frac{j\pi}{n+1}\right)}. \tag{15}$$

Let,

$$G'_{pp} = \frac{1}{\pi}\int_0^{\pi} e^{2\cos\theta}d\theta - \frac{1}{\pi}\int_0^{\pi}\cos\left(2p\theta\right)e^{2\cos\theta}d\theta$$

$$= I_0\left(2\right) - I_{2r(p)}\left(2\right), \tag{16}$$

where θ and $I_\gamma\left(z\right)$ are as before, and

$$r\left(p\right) = \begin{cases} p & \text{if } p \leqslant n/2 (n \text{ even}) \text{ or } p \leqslant (n+1)/2 (n \text{ odd}), \\ n-p+1 & \text{if } p > n/2 (n \text{ even}) \text{ or } p > (n+1)/2\ (n \text{ odd}). \end{cases}$$

The use of the term $r\left(p\right)$ is needed here because of the equivalence of the nodes v_i and v_{n-i+1} in a path. Here again $G_{pp}/G'_{pp} \to 1$ as $n \to \infty$ and by substitution we finally obtain the result. □

The previous result has some interesting implications. For instance, if we consider the two endpoints of a path we can see that $G_{n1} \to 0$ as $n \to \infty$, which implies that for very large paths the distance between the two endpoints is constant,

$$\xi_{n1} \to \sqrt{2\left[I_0\left(2\right) - I_2\left(2\right)\right]} \text{ as } n \to \infty. \tag{17}$$

The result can be extended to other pairs of nodes in the path by using results on banded Toeplitz matrices and on decay. However, we only want to give here the flavor that if a path $P_n\ (n \to \infty)$ is embedded into a Euclidean space using the communicability distance, the linear chain will fold in such a way that the two endpoints will remain at approximately the same distance.

Now, we continue with the analysis of the communicability distance in other simple graphs, starting by the cycle.

Lemma 3. *Let C_n be a cycle of n nodes labeled in clockwise order as $1, 2, \ldots, n$, and let $p \geqslant q$. Let,*

$$\xi'_{pq} = \sqrt{2\left[I_0\left(2\right) - I_{d_{pq}}\left(2\right)\right]}, \tag{18}$$

where d_{pq} is the shortest path distance between the two nodes. Then, the communicability distance between the nodes p and q in C_n tends to ξ'_{pq} as the size of the cycle increases, i.e., $\xi_{pq}/\xi'_{pq} \to 1$ as $n \to \infty$.

Proof. Using a geometric argument Spielman [20] has shown that the eigenvectors of the cycle graph are (notice that the eigenvectors of the adjacency and those of the Laplacian matrix found by Spielman are the same for the cycle): $\varphi_j\left(p\right) = \sin\left(2\pi jp/n\right)$ and $\varphi'_j\left(p\right) = \cos\left(2\pi jp/n\right)$, for $1 \leqslant j \leqslant r/2$ ($r = n$ if n is even and $r = n-1$ if n is odd). When n is even, $\boldsymbol{\varphi}_{n/2} = \mathbf{0}$ and we only have $\boldsymbol{\varphi}'_{n/2}$. The eigenvalues of the cycle graph are $\lambda_j = \cos\left(\frac{2\pi j}{n}\right)$. The reader surely noticed already that the adjacency matrix of a cycle is a circulant one and so any function of it. Then, every diagonal entry of $\exp\left(A\right)$ is equal to $\left[tr \exp\left(A\right)\right]/n$. Consequently, we set $p = 1$ and obtain

$$\begin{aligned} G_{pp} &= G_{11} = \frac{1}{n}\sum_{k=0}^{n/2}\cos^2\left(\frac{2\pi j}{n}\right)e^{2\cos\left(\frac{2\pi j}{n}\right)} + \frac{1}{n}\sum_{k=0}^{n/2}\sin^2\left(\frac{2\pi j}{n}\right)e^{2\cos\left(\frac{2\pi j}{n}\right)} \\ &= \frac{1}{n}\sum_{k=0}^{n/2}e^{2\cos\left(\frac{2\pi j}{n}\right)}\left\{\cos^2\left(\frac{2\pi j}{n}\right)+\sin^2\left(\frac{2\pi j}{n}\right)\right\} \\ &= \frac{1}{n}\sum_{k=0}^{n/2}e^{2\cos\left(\frac{2\pi j}{n}\right)}. \end{aligned} \tag{19}$$

Let,

$$\begin{aligned} G'_{pp} &= G'_{11} = \frac{2}{n}\int_0^{\pi} e^{2\cos\left(\frac{2\pi j}{n}\right)}dj = \frac{1}{\pi}\int_0^{\pi} e^{2\cos\theta}d\theta \\ &= I_0\left(2\right), \end{aligned} \tag{20}$$

where $\theta = 2\pi j/n$. As in the case of Lemma 2 we can easily see that $G_{pp}/G'_{pp} \to 1$ as $n \to \infty$.

Similarly, we can write the communicability function for any pair of nodes as

$$\begin{aligned} G_{pq} &= \frac{1}{n}\sum_{k=0}^{n/2}\cos\left(\frac{2\pi jp}{n}\right)\cos\left(\frac{2\pi jq}{n}\right)e^{2\cos\left(\frac{2\pi j}{n}\right)} + \frac{1}{n}\sum_{k=0}^{n/2}\sin\left(\frac{2\pi jp}{n}\right)\sin\left(\frac{2\pi jq}{n}\right)e^{2\cos\left(\frac{2\pi j}{n}\right)} \\ &= \frac{1}{n}\sum_{k=0}^{n/2}e^{2\cos\left(\frac{2\pi j}{n}\right)}\cos\left(\frac{2\pi j\left(p-q\right)}{n}\right). \end{aligned} \tag{21}$$

Let,

$$\begin{aligned} G'_{pq} &= \frac{1}{\pi}\int_0^{\pi}\cos\left[\theta\left(p-q\right)\right]e^{2\cos\theta}d\theta \\ &= I_{d_{pq}}\left(2\right), \end{aligned} \tag{22}$$

where $\theta = 2\pi j/n$, $I_\gamma\left(z\right)$ is the Bessel function of the first order [18] and $d_{p,q}$ is the shortest path distance between the two nodes. By similar arguments as before we have $G_{pq}/G'_{pq} \to 1$ as $n \to \infty$. Then, by substitution we obtain the final result. □

It can be seen that $I_{d_{pq}}\left(2\right) \to 0$ as $d_{pq} \to \infty$, which implies that the communicability distance tends to a constant for pairs of nodes separated at very large shortest path distance, $\xi_{pq} \to \sqrt{2I_0\left(2\right)} = 2.1352\ldots$ as $d_{pq} \to \infty$.

Lemma 4. *Let $S_{1,n-1}$ be the star graph of n nodes. The communicability distances between pairs of nodes in $S_{1,n-1}$ are given by*

$$\xi_{p1} = \sqrt{\frac{n}{n-1}\cosh\left(\sqrt{n-1}\right) - \frac{2}{\sqrt{n-1}}\sinh\left(\sqrt{n-1}\right) + \frac{n-2}{n-1}}, \quad p \neq 1, \tag{23}$$

$$\xi_{pq} = \sqrt{2}, \quad p \neq q \neq 1. \tag{24}$$

Proof. The eigenvalues of the star graph are $\sqrt{n-1}, 0, \ldots, 0, -\sqrt{n-1}$. The eigenvectors associated with the largest and smallest eigenvalues are, respectively [21]:

$$\varphi_1 = \left(1/\sqrt{2}\ \ 1/\sqrt{2\left(n-1\right)}\ \cdots\ 1/\sqrt{2\left(n-1\right)}\right), \tag{25}$$

$$\varphi_n = \left(-1/\sqrt{2}\ \ 1/\sqrt{2\left(n-1\right)}\ \cdots\ 1/\sqrt{2\left(n-1\right)}\right). \tag{26}$$

The other expressions that we need to know in order to find the formulae for the communicability among nodes in a star are:

$$\sum_{1<j<n} [\varphi_j(p)]^2 = 0, \quad p \neq 1, \tag{27}$$

$$\sum_{1<j<n} \varphi_j(p)\,\varphi_j(q) = -\frac{1}{n-1}, \quad p \neq q \neq 1, \tag{28}$$

which are derived from the orthonormality of the eigenvectors of the adjacency matrix. Then, the self-communicability of the nodes in $S_{1,n-1}$ are given by

$$G_{11} = \cosh\left(\sqrt{n-1}\right), \tag{29}$$

$$G_{pp} = \frac{1}{n-1}\left[\cosh\left(\sqrt{n-1}\right) + n - 2\right], \quad p \neq 1, \tag{30}$$

and the communicability between pairs of nodes in a star are

$$G_{p1} = \frac{1}{\sqrt{n-1}} \sinh\left(\sqrt{n-1}\right), \quad p \neq 1, \tag{31}$$

$$G_{pq} = \frac{1}{n-1}\left[\cosh\left(\sqrt{n-1}\right) - 1\right], \quad p \neq q \neq 1. \tag{32}$$

Thus, by substitution in (4) we obtain the expressions for the communicability distances between pairs of nodes in the star graph. □

Finally, we obtain the value for the communicability distance between a pair of nodes in a complete graph.

Lemma 5. *Let K_n be the complete graph of n nodes. The communicability distance between any pair of nodes is given by*

$$\xi_{pq} = \sqrt{2}e^{-1/2} = 0.85776\ldots. \tag{33}$$

Proof. The communicability between any pair of nodes in K_n is given by

$$G_{pq} = \frac{e^{n-1}}{n} + e^{-1}\sum_{j=2}^{n} \varphi_j(p)\varphi_j(q) = \frac{e^{n-1}}{n} - \frac{1}{ne} = \frac{e^n - 1}{ne}. \tag{34}$$

Similarly,

$$G_{pp} = \frac{e^{n-1}}{n} + \frac{n-1}{ne}, \tag{35}$$

which by substitution in (3) gives the final result. □

This result indicates that independently of the size of the complete graph the communicability distance between its nodes is always a constant. The value of the communicability distance between two nodes of a complete graph is relatively small, which indicates that when a simultaneous excitation arises in these nodes, the amount of excitation absorbed and transmitted is practically the same.

3. Communicability distance sum in graphs

We start by defining a graph invariant in a similar way as the Wiener [22] and Kirchhoff indices of graphs [7] are defined. We recall that the Wiener index is defined as $W(G) = \sum_{p<q} d_{pq}$ and the

Kirchhoff index as $Kf(G) = \sum_{p<q} \Omega_{pq}$. First, we construct the communicability distance matrix of a graph as follows. Let $\mathbf{s} = \begin{bmatrix} G_{11} & G_{22} & \cdots & G_{nn} \end{bmatrix}^T$ be a column vector of the self-communicabilities of every node in the graph. Let **M** be the matrix defined as:

$$\mathrm{M} = \mathbf{s}\mathbf{1}^T + \mathbf{1}\mathbf{s}^T - 2e^{\mathbf{A}}, \tag{36}$$

where **1** is a column vector of ones. Then, the communicability distance matrix of a graph, which is a Euclidean distance matrix, is given by

$$\mathbf{X}(G) = \sqrt[\circ]{\mathbf{M}}, \tag{37}$$

where $\sqrt[\circ]{\ }$ is the entrywise square root.

Definition 1. The communicability distance index of a simple connected graph $\Upsilon(G)$ is defined as:

$$\Upsilon(G) = \frac{1}{2}\mathbf{1}^T\mathbf{X}(G)\,\mathbf{1} = \frac{1}{2}\sum_{p,q}\xi_{pq}. \tag{38}$$

A small value of $\Upsilon(G)$ indicates that the nodes of the graph are close to each other in the sense of mutually 'feeling' their excitations. In such a way the index $\Upsilon(G)$ accounts for the global communicability 'packing' of the graph.

We have found the expressions for $\Upsilon(G)$ for the different types of graphs studied in the previous section. We start by studying the path, for which we plot the relationship between the communicability distance index and the number of nodes (see Fig. 1). As can be seen from the Fig. 1 there is a quadratic dependence between $\Upsilon(G)$ and n obeying the following equation:

$$\Upsilon(P_n) = 1.0676n^2 - 2.8725n + 2.4823, \tag{39}$$

with Pearson correlation coefficient $r > 0.9999$ and norm of residuals equal to 0.0008.

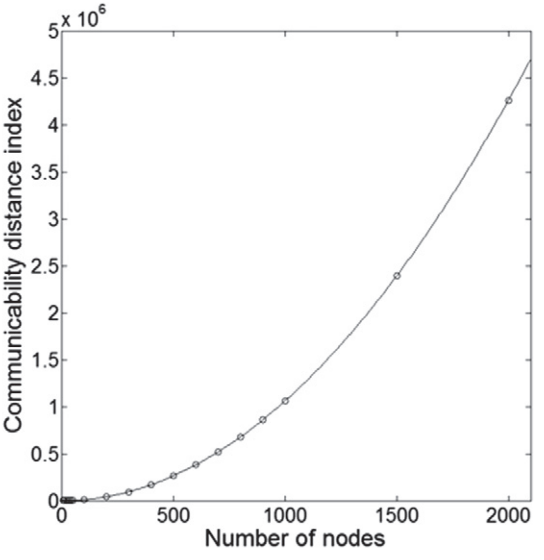

Fig. 1. Quadratic fit of the communicability distance index of paths with the number of nodes.

A different approach was followed for the cycle graphs. In this case we have that $I_{d_{pq}}(2) \to 0$ as $d_{pq} \to \infty$. In practice $I_{d_{pq}}(2)$ is almost zero for shortest-path distances larger than 6. Then, we can approximate the expression for $\Upsilon(G)$ by calculating the contributions of the terms $I_0(2) - I_{d_{pq}}(2)$ up to $d_{pq} = 6$, and for $d_{pq} > 6$ we can consider that $\xi_{pq} \simeq \sqrt{2I_0(2)}$. Let,

$$\Upsilon'(C_n) = n\left[(n-1)\frac{\sqrt{2I_0(2)}}{2} - 1.4443\right] = n\left[1.0676(n-1) - 1.4443\right]. \quad (40)$$

Then, $\Upsilon(G)/\Upsilon'(G) \to 1$ as $n \to \infty$. In fact, $\Upsilon(G)/\Upsilon'(G) = 1.0000$ for $n \geqslant 9$, which indicates that (40) is an excellent approximation.

The expressions for Υ of the star and complete graphs are given below:

$$\Upsilon(S_n) = (n-1)\left[\sqrt{\xi_{1,p}} + \frac{\sqrt{2}}{2}(n-2)\right], \quad (41)$$

$$\Upsilon(K_n) = n(n-1)\frac{\sqrt{2}e^{-1/5}}{2}. \quad (42)$$

4. Computational studies

We have calculated the $\Upsilon(G)$ index for all the 12,111 connected graphs with $n = 3$ to 8 nodes. We have found that among the graphs with n nodes the complete graph K_n always has the smallest value of $\Upsilon(G)$. Then, we have the following:

Conjecture 1. *Let $G_n \neq K_n$ be a simple connected graph with n nodes. Then,*

$$\Upsilon(G_n) > \Upsilon(K_n). \quad (43)$$

Among the trees of n nodes the star always has the smallest and the path has the largest value of $\Upsilon(G)$. Thus, we have the following:

Conjecture 2. *Let T_n be a tree with n nodes. Then,*

$$\Upsilon(S_n) \leqslant \Upsilon(T_n) \leqslant \Upsilon(P_n). \quad (44)$$

Let G_n be a simple connected graph with $n \leqslant 5$ nodes. Then $\Upsilon(G_n) \leqslant \Upsilon(P_n)$. If G_n is a graph with $n > 5$ nodes we have observed that the lollipop graph $L_{n-2,2}$ has the maximum value of $\Upsilon(G)$. The lollipop graph $L_{r,s}$ is formed by joining a complete graph K_r and the path P_s by a bridge (see Fig. 2). Thus we have the following:

Conjecture 3. *Let G_n be a simple connected graph with $n > 5$ nodes. Then,*

$$\Upsilon(K_n) \leqslant \Upsilon(G_n) \leqslant \Upsilon(L_{n-2,2}). \quad (45)$$

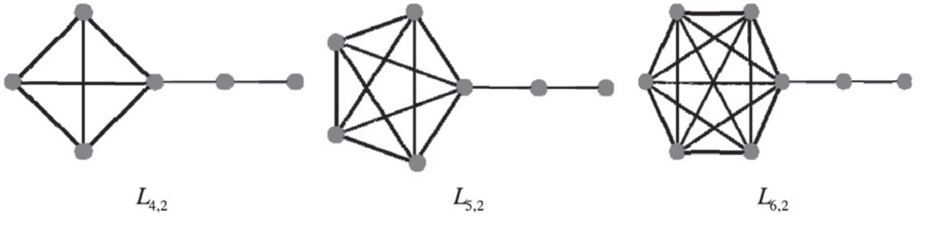

Fig. 2. The lollipop graphs that maximize $\Upsilon(G)$ for graphs with 6, 7 and 8 nodes, respectively.

In a lollipop graph $L_{n-2,2}$ the communicability distance between a pair of nodes in the K_{n-2} subgraph or between the two nodes in the P_2 subgraph is very small. However, the communicability distance between a node $v_p \in K_{n-2}$ and a node $v_q \in P_2$ is very large, indicating that when there is a simultaneous excitation at these two nodes most of the excitation is dissipated among the rest of the nodes of the graph.

In order to gain insights about the nature of the communicability distance index we have explored the empirical relationships between this index and the Wiener and Kirchhoff indices for all 11,117 connected graphs with 8 nodes. We start by illustrating the scatterplot of these indices for trees. Notice that for trees the Wiener and the Kirchhoff indices coincide. As can be seen in Fig. 3a the Wiener index and the communicability distance index display some kind of nonlinear relationship. Despite of this empirical relationship both indices are not trivially related to each other. For instance, there are three pairs of trees with 8 nodes that have the same value of the Wiener/Kirchhoff index, but which are differentiated by the communicability distance index (see Fig. 4). In general, $\Upsilon(G)$ is not related neither to the Wiener nor to the Kirchhoff indices as can be seen in Fig. 3b and c, where the lack of any empirical relationship between these indices is very clear.

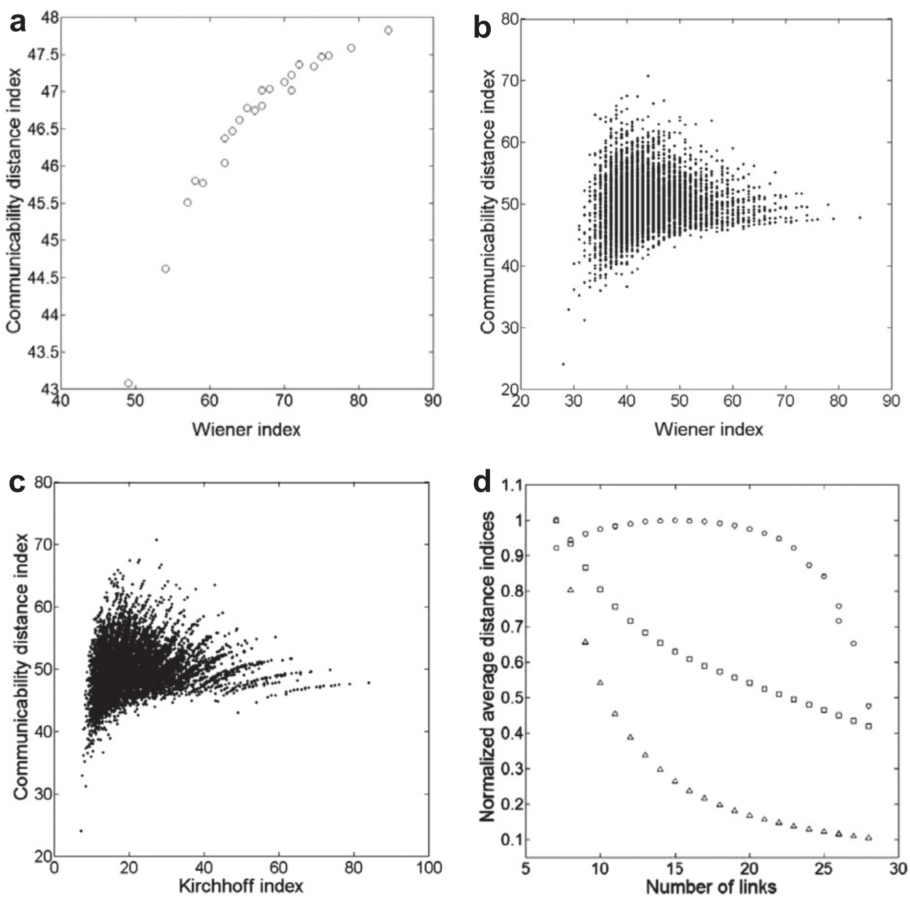

Fig. 3. Illustration of the empirical relationships between Wiener, Kirchhoff and the communicability distance indices for all trees with 8 nodes (a) and all connected graphs with 8 nodes (b and c). (d) Plot of normalized average values of the Kirchhoff index (triangles), the Wiener index (squares) and the communicability distance index (circles) versus the number of edges of connected graphs with 8 nodes.

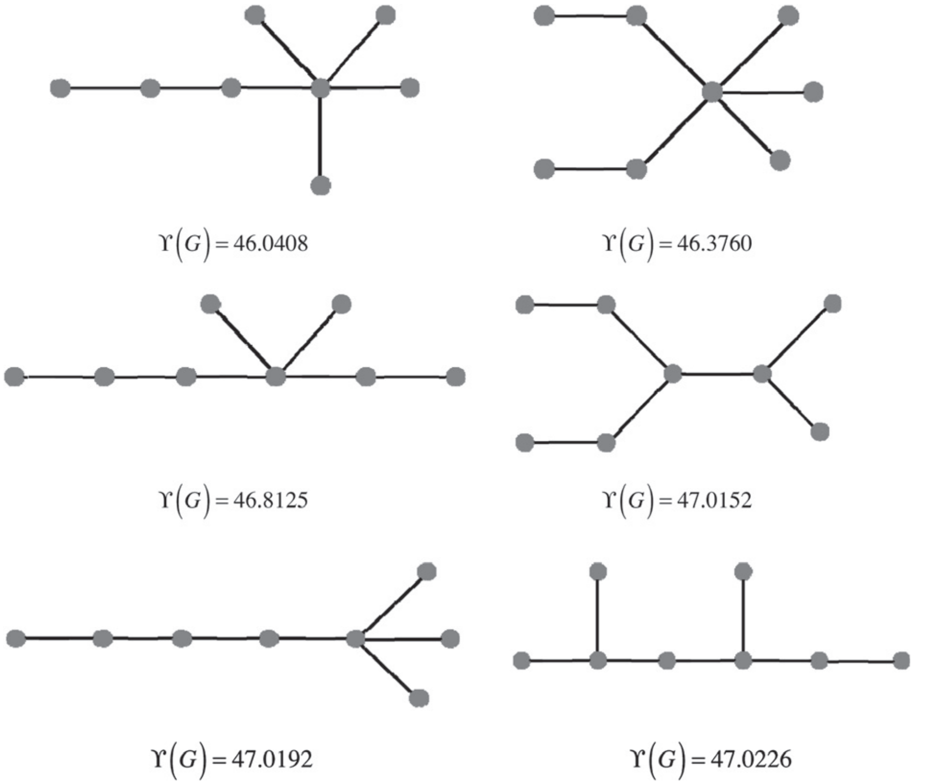

Fig. 4. Illustration of the three pairs of trees having the same value of the Wiener/Kirchhoff indices but different values of the communicability distance index. The pair at the top has $W(G) = Kf(G) = 62$, the pair in the middle has $W(G) = Kf(G) = 67$, and the pair at the bottom has $W(G) = Kf(G) = 71$.

Finally, we have studied how the three indices are affected by the density of edges in the graphs. We have calculated the average of the three indices for graphs having 8 nodes and m edges. For instance, we consider the average Wiener, Kirchhoff and communicability distance indices for all trees ($m = 7$), then for all monocycles, bicycles, etc. The results are illustrated in Fig. 3d. As can be seen both the Wiener and the Kirchhoff indices decrease monotonically as the number of edges increases. It means that for these two indices the trees display the largest average value and it decreases as the number of cycles in the graph increases. Interestingly, the communicability distance index displays a completely different behavior. Trees does not display the largest average value of this index but as the number of cycles increases the index starts to growth up to a maximum reached when $m = 15$. After this point the index decreases up to its minimal value obtained for the complete graph. This trend indicates a rather unique characteristic of the communicability distance. It depends on both the shortest path separation between the nodes and the 'cyclicity' around them. A qualitative analysis of this effect is as follows. Trees have the largest average path length but have no cyclicity. As soon as the number of edges increases the number of cycles also increases, which means that the cyclicity of the graph increases. However, the effect of increasing the number of edges also decreases the average path length of the graph. At certain point, here $m = 15$, the graphs have the largest possible cyclicity without reducing too much the average path length. This is the point where the average communicability distance index is maximum. This reasoning also explains why the lollipop graphs display the maximum value of $\Upsilon(G)$ among the graphs with $n > 5$ nodes. A quantitative analysis of these effects is out of the scope of this work and will be treated in details in a forthcoming paper dealing with more computational aspects of these indices. We finally recall that the average path length and the cyclicity of a graph are the two ingredients of the celebrated 'small-world' effect [23], which indicates that the communicability distance has an important role in explaining this effect observed in most of the real-world networks.

In summary, we have defined a Euclidean distance based on the communicability between a pair of nodes in a graph. Because the communicability between a pair of nodes has a clear physical interpretation [1–5] we foresee an important place for the new distance in the study of real-world systems.

Acknowledgements

The author thanks illuminating discussions with Prof. M. Saerens (Louvain). I am indebted to Prof. Michele Benzi (Emory) and Prof. Pavel Chebotarev (Moscow) for careful reading of the manuscript and very useful suggestions. Partial financial support from New Professor's Fund, University of Strathclyde and from the project *Mathematics of Large Technological Evolving Networks* (*MOLTEN*) supported by the Engineering and Physical Sciences Research Council and the Research Councils UK Digital Economy programme, with Grant Ref. EP/I016058/1 is also acknowledged.

References

[1] E. Estrada, N. Hatano, Communicability in complex networks, Phys. Rev. E 77 (2008) 036111
[2] E. Estrada, J.A. Rodríguez-Velázquez, Subgraph centrality in complex networks, Phys. Rev. E 71 (2005) 056103
[3] E. Estrada, Characterization of 3D molecular structure, Chem. Phys. Lett. 319 (2000) 713–718.
[4] E. Estrada, N. Hatano, Statistical–mechanical approach to subgraph centrality in complex networks, Chem. Phys. Lett. 439 (2007) 247–251.
[5] E. Estrada, N. Hatano, M. Benzi, The physics of communicability in complex networks, Phys. Rep., in press, arXiv:1109.2950v1 doi:10.1016/j.physrep.2012.01.006.
[6] F. Buckley, F. Harary, Distance in Graphs, Addison-Wesley, Redwood City, CA, 1990.
[7] D.J. Klein, M. Randić, Resistance distance, J. Math. Chem. 12 (1993) 81–95.
[8] A. Ghosh, S. Boyd, A. Saberi, Minimizing effective resistance of a graph, SIAM Rev. 50 (2008) 37–66.
[9] X. Xiao, I. Gutman, Resistance distance and Laplacian spectrum, Theor. Chem. Acc. 110 (2003) 284–289.
[10] P. Chebotarev, E. Shamis, The forest metric for graph vertices, Electron. Notes Discrete Math. 11 (2002) 98–107.
[11] P. Chebotarev, A class of graph-geodetic distances generalizing the shortest-path and the resistance distances, Discrete Appl. Math. 159 (2011) 295–302.
[12] P. Chebotarev, The walk distances in graphs, arXiv:1103.2059v6.
[13] M.M. Deza, E. Deza, Encyclopedia of Distance, Springer-Verlag, Berlin, 2009.
[14] J.C. Gower, Euclidean distance geometry, Math. Sci. 7 (1982) 1–14.
[15] J.C. Gower, Properties of Euclidean and non-Euclidean distance matrices, Linear Algebra Appl. 67 (1985) 81–97.
[16] A.Y. Alfakih, On the nullspace, the rangespace and the characteristic polynomial of Euclidean distance matrices, Linear Algebra Appl. 416 (2006) 348–354.
[17] J. Datorro, Convex Optimization and Euclidean Distance Geometry, Meboo Pub., Palo Alto, CA, 2011.
[18] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, Dover, New York, 1965.
[19] L. Collatz, U. Sinogowitz, Spektren endlicher Grafen, Abh. Math. Sem. Univ. Hamburg 21 (1957) 63–77.
[20] D. Spielman, Spectral Graph Theory. The Laplacian. Available from: <http://www.cs.yale.edu/homes/spielman/561/lect02-09.pdf>.
[21] G. Rücker, C. Rücker, I. Gutman, On kites, comets and stars. Sum of eigenvector coefficients in (molecular) graphs, Z. Naturforsch 57a (2002) 143–153.
[22] H. Wiener, Structural determination of paraffin boiling points, J. Amer. Chem. Soc. 69 (1947) 17–20.
[23] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 409–410.