



HAL
open science

Communicability cosine distance: similarity and symmetry in graphs/networks

Ernesto Estrada

► **To cite this version:**

Ernesto Estrada. Communicability cosine distance: similarity and symmetry in graphs/networks. 2023. hal-04169459

HAL Id: hal-04169459

<https://hal.science/hal-04169459v1>

Preprint submitted on 24 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Communicability cosine distance: similarity and symmetry in graphs/networks

Ernesto Estrada

the date of receipt and acceptance should be inserted later

Abstract A distance based on the exponential kernel of the adjacency matrix of a graph and representing how well two vertices connect to each other in a graph is defined and studied. This communicability cosine distance (CCD) is a Euclidean spherical distance accounting for the cosine of the angles spanned by the position vectors of the graph vertices in this space. The Euclidean distance matrix (EDM) of CCD is used to quantify the similarity between vertices in graphs and networks as well as to define a local vertex invariant—a closeness centrality measure, which discriminates very well vertices in small graphs. It allows to distinguish all nonidentical vertices, also characterizing all identity (asymmetric) graphs—those having only the identity automorphism—among all connected graphs of up to 9 vertices. It also characterizes several other classes of identity graphs. We also study real-world networks in terms of both the discriminating power of the new centrality on their vertices as well as in ranking their vertices. We analyze some dictionary networks as well as the network of copurchasing of political books, remarking some of the main advantages of the new approaches studied here.

Keywords: communicability; matrix functions; graph automorphism; centrality measures; graph symmetry; vertex similarity

Mathematics Subject Classification: 05C12; 05C50; 05C82; 05E05; 05C69

1 Introduction

In the classic book of Wasserman and Faust [1] it is stated that “one of the primary uses of graph theory in social network analysis is the identification of the “most important” actors in a social network”. Such identification is mainly done by means of the so-called “centrality measures”. A vertex centrality measure is a local vertex invariant, i.e., a property of vertices that is equal for pairs of vertices interchangeable by an automorphism [2]. Therefore, the vertex degree is a centrality measure, i.e., the more connections a vertex has, the most important it is. Although the pioneering work of

Institute for Cross-Disciplinary Physics and Complex Systems (IFISC) CSIC-UIB, Palma de Mallorca, Spain. E-mail: estrada@ifisc.uib-csic.es

Bavelas [3] about centrality were based on counting walks in the graph, the influential works of Freeman [4, 5] about “closeness” and “betweenness” centrality put a lot of emphasis on counting geodesics—shortest paths—in the graphs. The shortest path (SP) between two vertices is the shortest (in terms of number of edges) of all sequences of distinct vertices/edges (paths) connecting the two vertices. Differently, a walk allows for the repetition of both vertices and edges. The closeness centrality (CC) accounts for how close in terms of SP a vertex is in relation to the rest of the vertices in the graph. The betweenness centrality (BC) refers to the importance that a vertex has in the SP communication between other pairs of vertices (see [6] for definitions and properties of these and other centrality measures).

What it should mean, for instance, that the peaks of two mountains are separated at a distance x ? This information is very valuable for an eagle who can fly from one peak to the other. For a mountaineer, this information is useless because she cannot fly, but has to go down from the first peak and hike the second one. The same happens if we say that the average SP distance in a network is y (see [7]). If “items” are not traveling through the shortest paths—which can only be the case when the sender has complete information about the topology of the network—the information about the SP is useless. It is not surprising, however, that such measures like the CC based on SP brings some information about dynamical processes which do not take place through the SP but using diffusive patterns of spreading. These are the cases, for instance, of epidemic spreading or synchronization of couple oscillators, which are well-known diffusive processes. What happen is similar to the fact that the mountaineer can guess how much effort it will cost to go down the first mountain and hike the second one by using the distance between the two peaks. CC based on SP can “guess” how a diffusive process takes place (see for instance [8]), but it does not provide a precise information about such processes because it is out of its designed scope.

Nowadays centrality measures are used beyond social network analysis in a broad spectrum of computational analysis of networked systems. They include, for instance, use in dynamics on networks (synchronization, consensus, epidemics, information spreading), pattern recognition, community detection [9, 10], dimensionality reduction [11] and isomorphism detection [12]. The last two topics are of particular interest for the current work. The main idea of dimensionality reduction is to “reduce” a large-scale network to a smaller one while preserving some of its global structural characteristics. Chen et al. [11] proposed to use centrality measures for this purpose, such that by removing the least central vertices they were reducing the size of the networks while observing that certain structural features were preserved. The second application is that of determining whether two graphs are isomorphic, for which there exist no deterministic polynomial-time algorithm. The strategy is to rank the vertices of two test graphs using a given centrality [12]. If they are not the same, then the two graphs are not isomorphic. If they have the same ranked sequences, then they are declared to be potentially isomorphic and confirmed through additional heuristics. Both applications of centrality measures, dimensionality reduction and graph isomorphism detection, are heavily dependent on the discriminant power of the centrality measures used. For instance consider a regular graph. In this case, the eigenvector centrality (EC)—the entries of the eigenvector associated to the spectral radius of the adjacency matrix (see [6])—is the same for every node. Therefore, there is no chance of ranking the nodes to start a pruning based on the least central ones. If you are detecting whether two k -regular graphs of degree k are isomorphic or not using the EC you will end up always by using an additional heuristics because both graphs have the same list of EC. The dis-

criminant power of a centrality measure in general refers to the capacity of the index to differentiate pairs of vertices [13]. A better definition should be given in term of the capacity of a centrality index to differentiate nonidentical pairs of vertices, where identical vertices are those which can be swapped by an automorphism.

The problem of graph automorphism is intimately related to the symmetry of graphs. This problem has been treated in the mathematical literature using several approaches [16, 17]. More recently, the problem of analyzing the size of the automorphism group in large real-world networks has been explored [18–22]. The size of the automorphism group refers to the number of symmetric operations which swap the vertices of the graphs leaving unaltered its adjacency matrix. Then, it is clear that a complete graph on four vertices K_4 is more symmetric than a cycle graph of the same size C_4 . While the first has $|\text{Aut}(K_4)| = 24$ —there are $n!$ different symmetries in K_n , the second has $|\text{Aut}(C_4)| = 8$, n rotational and n reflection symmetries in C_n . However, from a vertex point of view it is true that both graphs have exactly the same number of identical pairs of vertices, which is equal to four. There are many more ways of swapping the vertices in K_n than in C_n but what an “inspector” of these graphs “sees” is that both are formed by four identical vertices. This is the important problem of vertex similarity—more appropriately vertex identity [24–28]. Symmetry groups has also been used as a proxy for the calculation of SP distance invariants in graphs [23].

Here, I will propose a distance metric between vertices in a simple graph. I will prove that this distance, which appears naturally from the way in which diffusive objects navigate a graph, is Euclidean and spherical, i.e., induces an embedding of the graph in a Euclidean hypersphere. I will then propose to use the Euclidean distance matrix (EDM) of communicability cosine distances (CCD) as a quantitative proxy to detect similarity between vertices and show some problems arising with a popular existing measure. Such drawbacks are corrected by the new method proposed here. Further I will defined and study a measure of closeness centrality based on CCD. This measure differentiates very well vertices in small graphs, allowing to distinguish all nonidentical vertices in connected graphs of up to 9 vertices. I also study several identity (asymmetric) graphs with this measure, all of whose vertices are differentiated by the new closeness centrality. Real-world networks are also studied for both the discriminating power of the new centrality on their vertices as well as for ranking their vertices. Several of the advantages of the communicability closeness centrality proposed here are analyzed along the paper.

2 Preliminaries

Here we consider simple (finite, undirected, unweighted and connected) graphs $G = (V, E)$ with $\#V = n$ and $\#E = m$. As usual, A denotes the adjacency matrix of the graph G , which is obviously symmetric for the kind of graphs we study here. We consider that $A = U\Lambda U^T$ where Λ is a diagonal matrix of eigenvalues of A and U is an orthogonal matrix of eigenvectors.

A walk of length l in a graph is a sequence of (not necessarily different) vertices $v_1, v_2, \dots, v_l, v_{l+1}$ such that for each $i = 1, 2, \dots, l$ there is an edge from v_i to v_{i+1} . The walk is known as closed if $v_{l+1} = v_1$. The number of walks of length l between v_p and v_q is given by $\left(A^l\right)_{pq}$. Hereafter we will designate the vertices directly by the

subindex, such as p and q . A path is a walk in which there is no repetition neither of vertices nor of edges. We also denote by u an all-ones vector.

Let us now consider a few definitions related to the communicability function of networks [29, 30].

Definition 1 The communicability function between two nodes $v, w \in V$ of G is given by

$$G_{vw} = (\exp(\beta A))_{vw}, \quad (2.1)$$

where $\exp(\beta A) = I + \beta A + \frac{(\beta A)^2}{2!} + \dots + \frac{(\beta A)^k}{k!} + \dots$ is the matrix exponential of A and β is an empirical parameter which may be set to one for the sake of simplicity.

Obviously, due to the spectral decomposition of the adjacency matrix we have

$$G_{vw} = \sum_{j=1}^n \psi_{jv} \psi_{jw} \exp(\beta \lambda_j), \quad (2.2)$$

where ψ_{jv} is the v th entry of the eigenvector associated with the j th eigenvalue λ_j of A .

Definition 2 The term $G_{vv} = (\exp(\beta A))_{vv}$ is the subgraph centrality [31] of the node v and it accounts for the importance of a node in terms of its participation in all subgraphs of the network giving more importance to the smaller than to the larger ones.

Definition 3 The terms [32, 33]

$$\xi_{vw} := G_{vv} + G_{ww} - 2G_{vw}, \quad (2.3)$$

and

$$\zeta_{vw} := \frac{G_{vw}}{\sqrt{G_{vv}G_{ww}}}, \quad (2.4)$$

account for the proximity between the nodes v and w in terms of the way in which they transmit ‘‘information’’ from each other. If ξ_{vw} or ζ_{vw} are small it means that most of the information sent from v (resp. w) to w (resp. to v) arrives to its destination, in contrast to the information which is returned to the origin.

Lemma 1 *The terms ξ_{vw} and ζ_{vw} are a squared Euclidean distance between the vertices v and w and cosine of the angle between the position vectors of both vertices in a Euclidean space, respectively.*

Proof First let us write

$$\begin{aligned}
G &= e^A \\
&= Ue^{\Lambda}U^T \\
&= Ue^{\Lambda/2}e^{\Lambda/2}U^T \\
&= \left(e^{\Lambda/2}U^T\right)^T \left(e^{\Lambda/2}U^T\right) \\
&= X^T X.
\end{aligned} \tag{2.5}$$

Therefore,

$$G_{vw} = x_v \cdot x_w, \tag{2.6}$$

where $x_i = e^{\Lambda/2}\varphi_i$ represents the position vector of the vertex i in the Euclidean space induced by the communicability and φ_i is the i th column of U^T . Thus, we have

$$\begin{aligned}
\xi_{vw} &= x_v \cdot x_v + x_w \cdot x_w - 2x_v \cdot x_w \\
&= \|x_v - x_w\|^2,
\end{aligned} \tag{2.7}$$

and

$$\zeta_{vw} = \frac{x_v \cdot x_w}{\|x_v\| \|x_w\|} = \cos \theta_{vw}. \tag{2.8}$$

□

Another known Euclidean distance between the vertices of a graph is the so-called resistance distance, which is formally defined below.

Definition 4 Let $L = K - A$ be the Laplacian matrix of a graph, where K is the diagonal matrix of vertex degree and A its adjacency matrix. Then,

$$\Omega_{vw} := L_{vv}^+ + L_{ww}^+ - 2L_{vw}^+, \tag{2.9}$$

is the resistance distance [34] between the vertices v and w in the graph, where L^+ is the Moore-Penrose pseudoinverse of the Laplacian matrix.

Definition 5 Let d_{vw} and Ω_{vw} be the shortest-path and resistance distance between the vertices v and w in a graph. Then,

$$CC_v = \left(\sum_{w=1}^n d_{vw} \right)^{-1}, \tag{2.10}$$

and

$$RCC_v = \left(\sum_{w=1}^n \Omega_{vw} \right)^{-1}, \tag{2.11}$$

are the closeness [3] (see also [1, 6]) and resistance-closeness [35] centrality of the vertex v , respectively.

Other definitions needed in the current work are the following.

Definition 6 Let $D = (d_{ij})$ be an $n \times n$ matrix. Then, D is called Euclidean distance matrix (EDM) if there exist points p^1, p^2, \dots, p^n in some Euclidean space \mathbb{R}^r , such that

$$d_{ij} = \left\| p^i - p^j \right\|^2, \quad (2.12)$$

for all $i, j = 1, \dots, n$, where $\| \cdot \|$ denotes Euclidean norm. An EDM D is called spherical EDM or circum-EDM if the points p^1, p^2, \dots, p^n that generate D lie on a hypersphere [36, 37]. The matrix having entries given by ξ_{vw} is a spherical EDM [39].

Let us now consider some definitions on graph automorphism which will be used in this work (see [2]).

Definition 7 Let $G = (V, E)$ be a simple graph. A permutation α of V is an automorphism of G if

$$\{v, w\} \in E \iff \{\alpha(v), \alpha(w)\} \in E,$$

for all $v, w \in V$.

Definition 8 The set of all automorphisms of G , together with the operation of composition of functions, forms a subgroup of the symmetric group of V , which is called the automorphism group of G , and it is denoted by $\text{Aut}(G)$.

Definition 9 Two vertices v and w in a graph G are identical (usually called similar [24] in graph theory) if some automorphism of G maps v onto w . If v and w are identical vertices in G then the graphs $G - v$ and $G - w$, obtained by deleting v and w , respectively, from G , are isomorphic. However, this is not true the other way around because it is possible for $G - v$ and $G - w$ to be isomorphic even when v and w are not similar. In this case the nodes are pseudoidentical (usually pseudosimilar [25, 26] in graph theory).

Definition 10 A graph in which there is no pair of identical vertices is known as asymmetric or identity graph [27]. Let ι be the identity of any group of permutations. Then, the graph G is asymmetric if ι is its only automorphism.

3 Communicability cosine distance

The communicability distance ξ_{vw} between two vertices v and w , previously defined and studied, is an unbounded magnitude. On the contrary, the communicability angle θ_{vw} is always bounded between 0° and 90° for unweighted graphs [33]. This boundedness may represent an interesting characteristic for comparing properties within and between graphs. Therefore, here I propose a transformation of the communicability angle into a communicability cosine distance which is bounded and display a few mathematically interesting properties. Let me start with the following.

Definition 11 Let $\cos \theta_{vw}$ be the cosine of the communicability angle. Then

$$D_{vw} = 2 - 2 \cos \theta_{vw}, \quad (3.1)$$

is the squared communicability cosine distance (CCD) between the nodes v and w in G .

Definition 12 The CCD matrix D is the square, symmetric matrix whose v, w entry is D_{vw} .

Now, I will prove some of the mathematical properties of the CCD.

Lemma 2 *The CCD is a squared Euclidean distance between the corresponding pair of nodes in G .*

Proof Let us write

$$D_{vw} = 2 - 2 \frac{x_v \cdot x_w}{\|x_v\| \|x_w\|}, \quad (3.2)$$

which can be expressed as

$$\begin{aligned} D_{vw} &= \frac{x_v \cdot x_v}{\|x_v\|^2} + \frac{x_w \cdot x_w}{\|x_w\|^2} - 2 \frac{x_v \cdot x_w}{\|x_v\| \|x_w\|} \\ &= \left\| \frac{x_v}{\|x_v\|} - \frac{x_w}{\|x_w\|} \right\|^2, \end{aligned} \quad (3.3)$$

which proves the result. \square

Theorem 1 *The CCD matrix is a spherical EDM.*

Proof First, let write the matrix D as follows:

$$D = 2 \left(uu^T - S^{-1/2} e^A S^{-1/2} \right), \quad (3.4)$$

where S is the diagonal matrix with diagonal entries $S_{vv} = \left(e^A \right)_{vv}$.

To prove that D is circum-Euclidean it is needed to prove that D^{-1} exists and $u^T D^{-1} u > 0$. Then let me find D^{-1} using the Sherman-Morrison formula:

$$\frac{1}{2} D^{-1} = -S^{1/2} e^{-A} S^{1/2} - \frac{S^{1/2} e^{-A} S^{1/2} uu^T S^{1/2} e^{-A} S^{1/2}}{1 - u^T S^{1/2} e^{-A} S^{1/2} u}. \quad (3.5)$$

To prove that D^{-1} exists it is only needed to prove that $u^T S^{1/2} e^{-A} S^{1/2} u \neq 1$. Then, let $J = uu^T$ and write

$$\begin{aligned} a &= u^T S^{1/2} e^{-A} S^{1/2} u \\ &= \text{Tr} \left(S^{1/2} e^{-A} S^{1/2} J \right) \\ &= \text{Tr} \left(e^{-A} S^{1/2} J S^{1/2} \right) \\ &= \text{Tr} \left(e^{-A} Z \right), \end{aligned} \quad (3.6)$$

where $Z = S^{1/2}JS^{1/2}$.

Now, I will use Ruhe's trace inequality [38] to find

$$\begin{aligned} \text{Tr} \left(e^{-A}Z \right) &\geq \sum_{j=1}^n \lambda_j(Z) \lambda_{n-j+1} \left(e^{-A} \right) \\ &= \lambda_1(Z) \lambda_n \left(e^{-A} \right), \end{aligned} \quad (3.7)$$

where $\lambda_j(Z) = \{\lambda_1(Z), 0, \dots, 0\}$ are the eigenvalues of Z . It is straightforward to realize that $\lambda_1(Z) = EE(G) = \text{Tr} \left(e^A \right)$, which is known as the Estrada index of the graph (see [40]), and $\lambda_n \left(e^{-A} \right) = e^{-\lambda_1(A)}$, where $\lambda_1(A)$ is the spectral radius of A . Therefore,

$$\begin{aligned} a = \text{Tr} \left(e^{-A}Z \right) &\geq e^{-\lambda_1(A)} \text{Tr} \left(e^A \right) = e^{-\lambda_1(A)} \sum_{j=1}^n e^{\lambda_j(A)} \\ &= \sum_{j=1}^n e^{\lambda_j(A) - \lambda_1(A)} = 1 + \sum_{j=2}^n e^{\lambda_j(A) - \lambda_1(A)} > 1, \end{aligned} \quad (3.8)$$

with the last inequality due to the fact that the graph is not trivial.

Now, let me prove that $u^T D^{-1}u > 0$, for which I will start by writing

$$\begin{aligned} \frac{1}{2}u^T D^{-1}u &= -u^T S^{1/2}e^{-A}S^{1/2}u - \frac{u^T S^{1/2}e^{-A}S^{1/2}ee^T S^{1/2}e^{-A}S^{1/2}u}{1 - u^T S^{1/2}e^{-A}S^{1/2}u}. \\ &= -a - \frac{a^2}{1-a} = \frac{1}{a-1} > 0, \end{aligned} \quad (3.9)$$

where the last inequality is due to the fact that $a > 1$ as proved before. \square

Remark 1 The previous result indicates that the CCD induces an embedding of a graph into a Euclidean n -dimensional sphere of radius (see [39]):

$$R = \sqrt{\frac{1}{2}(u^T D^{-1}u)^{-1}} = \sqrt{a-1}. \quad (3.10)$$

In a recent paper [41] I have proved that every circum-EDM is the effective resistance matrix of a graph with appropriate edge weights.

4 Communicability cosine distance as vertex similarity

It is obvious that the EDM of communicability cosines distances represents a dissimilarity matrix where the nondiagonal entries represent the dissimilarity between the corresponding pair of nodes. Such dissimilarity accounts for how well communicated a pair of nodes are in the network based on the ratio of the number of weighted walks connecting them to the number of weighted walks that start and end in the corresponding nodes. Here, the weight is given by the reverse of the factorial of the length of the walk.

The use of different graph measures to account for vertex similarities is not new. In fact, the use of the so-called structural equivalence, where two vertices sharing many of the same network neighbors are considered structurally equivalent, dates back to the 1970's [42]. Further developments include the regular equivalence, in which vertices are said to be similar if they are connected to other vertices that are themselves similar [43, 44]. More recent approaches include the use of algebraic methods, such as the method proposed by Blondell et al. [45] as a generalization of Kleinberg's "hub and authority" method [46], and the method proposed by Leicht, Holme and Newman (LHN) [47] on the basis of the so-called Katz centrality index [48]. In this last approach two vertices are similar if their immediate neighbors in the network are themselves similar, which leads to a self-consistent matrix formulation of similarity that can be evaluated in an iterative way using only a knowledge of the adjacency matrix of the network. LHN [47] tested their measure, among other networks, for extracting sensible synonyms to words from a network representing the structure of Roget's Thesaurus [49], similarly to what Blondell et al. [45] have done before. In general, the authors considered that this new measure is capable of extracting useful information about vertex similarity based on network topology and that it displays some advantages in relation with previously defined measures. Therefore, I will concentrate here in the comparison between the CCD dissimilarity and that of LHN. More formally, the similarity matrix proposed by LHN [47] is as follows.

Definition 13 Let G be a graph with adjacency matrix A and let K be the diagonal matrix of vertex degrees. The LHN similarity between two nodes v and w is given by the v, w entry of the following matrix

$$\mathcal{N} = 2m\lambda_1 K^{-1} \left(I - \frac{\alpha}{\lambda_1} A \right)^{-1} K^{-1}, \quad (4.1)$$

where m is the number of edges, λ_1 is the spectral radius of A , I is the identity matrix and $0 < \alpha < 1$ is an empirical parameter.

LHN [47] performed a brief analysis of the parameter α in their paper determining empirically that a value of $\alpha = 0.97$ was the most appropriated for the analysis of vertex similarities in networks. Because $0 \leq D_{vw} \leq 2$, I will consider here $\mathcal{D}_{vw} = \frac{1}{2} D_{vw}$, which is then bounded as $0 \leq \mathcal{D}_{vw} \leq 1$ as required by any dissimilarity metric. I then applied here the two measures \mathcal{D}_{vw} and \mathcal{N}_{vw} to the analysis of similarities between words in the Roget Thesaurus network. It should be noticed from the eq. (4.1) that the diagonal entries of \mathcal{N} are not one, like it is common in all similarity matrices. However, it should be not a problem if $\mathcal{N}_{vv} > \mathcal{N}_{vw}$ for all $w \in V$. That is, a vertex is most similar to itself than with any other vertex in the graph. This is indeed observed for all the vertices in the network of the Roget Thesaurus. For the comparison of the two approaches I have considered the same words analyzed by LHN [47] and the results are displayed in Table 1.

alarm		hell	
\mathcal{N}_{vw}	\mathcal{D}_{vw}	\mathcal{N}_{vw}	\mathcal{D}_{vw}
warning	warning	heaven	pain
danger	omen	pain	dejection
omen	indication	discontent	heaven
indication	prediction	adversity	discontent
caution	threat	pleasure	physical pain
mean		water	
\mathcal{N}_{vw}	\mathcal{D}_{vw}	\mathcal{N}_{vw}	\mathcal{D}_{vw}
compromise	imperfection	plunge	plunge
middle	unimportance	moisture	insertion
generality	smallness	air	moisture
mid-course	inferiority	insertion	river
compensation	weakness	river	egress

Table 1: Top five most similar words found by the LHN similarity measure, \mathcal{M}_{vw} , and by the cosine communicability similarity, \mathcal{N}_{vw} , for the words “alarm”, “hell”, “mean” and “water”.

As can be seen in Table 1 there are some differences between the ranking of words performed by both approaches, but in general they are qualitatively very similar. For instance, for “alarm” both methods identify “warning” as the most similar one and then rank “omen” and “indication” in the top five. For “hell”, LHN identifies “heaven” as the top similar word while the new method identifies “pain” as the top one, which is then ranked as the second most similar by LHN. For “mean” the words in the top five according to both methods differ, but they are closely related words, which is understood by the very meanings of the words “mean”. Finally, there is coincidence in the top-ranked similar word for “water”, in which both methods coincide in “plunge”, and then coincide in identifying “moisture”, “insertion” and “river” in the top five.

However, an important difficulty can be foreseen for the general use of the LHN method [47] in real-world networks. The analysis of the formula (4.1) indicates that if a network has a very large spectral radius of the adjacency matrix, i.e., $\lambda_1 \gg 1$, then $\alpha/\lambda_1 \rightarrow 0$, which makes that the influence of the whole topology of the network, accounted for by the adjacency matrix, disappears from the similarity matrix. That is, in a network in which λ_1 is relatively large, it is expected that $\mathcal{N}_{vw} \sim \frac{2m\lambda_1}{k_v k_w}$, where k_v is the degree of the vertex v . That is, in this case the similarity between the two vertices will depend only on their respective degrees. The Roget Thesaurus network has $\lambda_1 \approx 12.027$, which according to the results previously obtained, still allows for capturing the influence of the topology of the network in the LHN formula. Let us then consider another network, for which $\lambda_1 \approx 44.303$. This network is a representation of the Online Dictionary for Library and Information Science by Joan M. Reitz [50]. That is, a dictionary which is specialized in library and information science.

The first curious result obtained with the LHN for this network is the following. Now, it is not always true that $\mathcal{N}_{vv} > \mathcal{N}_{vw}$. That is, there are some vertices for which there are other vertices more similar to them than themselves. This is certainly weird! For instance, for “homepage”, LHN finds that there are 22 other words more similar to it than the proper word “homepage”. As another example, LHN finds that the words “data” and “queue” are more similar than “data” to itself. For the word “book” there are 1682 words more similar to it than the word “book” itself, and for the word “work” there are 1378 words in the same situation. Why is this happening? Simply because in

this case λ_1 is sufficiently large as for $\alpha/\lambda_1 \rightarrow 0$, implying that $\mathcal{N}_{vw} \sim \frac{2m\lambda_1}{k_v k_w}$. Then, if the vertex v has a relatively large degree, it will have more similarity with those vertices w displaying a very low degree, e.g., $k_w = 1$. And this is exactly what happens for words like “work” ($k = 380$), “text” ($k = 246$) or “homepage” ($k = 150$), just to give three examples. In Table 2 I give the top seven most similar words according to LHN and to the CCD dissimilarity.

work		text	
\mathcal{N}_{vw}	\mathcal{D}_{vw}	\mathcal{N}_{vw}	\mathcal{D}_{vw}
juvenalia	original	bibliotaph	composition
potboiler	direct edition	book cradle	note
bibliotaph	national bibliography	bookrest	reading level
book cradle	series title page	imbrication	print
bookrest	annotated edition	lectionary	galley proof
imbrication	royalty	polaire	first edition
lectionary	subtitle	architectural rendering	extent
homepage		book	
\mathcal{N}_{vw}	\mathcal{M}_{vw}	\mathcal{N}_{vw}	\mathcal{M}_{vw}
DEMCO	Medical Library Association	bibliotaph	curiosa
Highsmith Inc.	Freedom to Read Foundation	book cradle	easel
Brodart	Librarian of Congress	bookrest	demand publishing
bibliotaph	library of science	imbrication	primer
book cradle	standing order	lectionary	book collecting
bookrest	curator	polaire	prepub
imbrication	National Film Registry	book light	remainders

Table 2: Top most similar words in ODLIS to the words “work”, “text”, “homepage” and “book” according to LHN method and the current approach. The words in bold correspond to unrelated words that LHN identifies simply because they have degree one and are connected to a hub in the ODLIS network.

As can be seen in Table 2 there are five words which are repeated among the top seven found as the most similar one for “work”, “text” and “homepage”. These words, according to their definitions in ODLIS, are completely unrelated to the target words as can be seen in Table 3. What happens is that these five words have degree one in the ODLIS network. Then, they are found as “most similar” simply because of the fact that the similarity is given here by $\mathcal{N}_{vw} \sim \frac{2m\lambda_1}{k_v k_w}$, such that low degree of vertex w will increase its similarity with vertex v . The fact that the words “juvenalia” and “potboiler” are among the top most similar ones with “work” is due to the fact that they have degree one, and they are connected to “work” in the network. The same happens for “DEMCO”, “Highsmith Inc.” and “Brodart” in relation with “homepage”. However,

because the word “text” has no word of degree one connected to it, LHN selects any word with degree one as the most similar to it.

word	ODLIS definition
bibliotaph	A person who hoards books and hides them from others, even to the extent of keeping them under lock and key.
book cradle	A low stand or rack, usually made of wood, metal, or plastic, designed to display a book open at an angle, rather than flat, to minimize strain on the spine when the volume is exhibited, mounted, studied, microfilmed, or scanned (see this soft-surfaced example).
bookrest	A portable device similar to the music rest on a piano, designed to be placed on a desk or table to hold a book at an angle convenient for reading.
imbrication	In the book arts, a decorative pattern designed to give the impression of overlapping scales, tiles, shingles, leaves, etc.
lectionary	A liturgical book containing lessons and selections from the Bible, also indicating the sequence in which they are to be read by the congregation in services throughout the year.
polaire	A sturdily made leather satchel or case used by monks, scribes, and other literate persons of the medieval period for transporting manuscript books.
architectural rendering	A pictorial representation of a building or other structure, usually from an angle showing the front or main entrance, created by the architect or an architectural firm to give an accurate, if somewhat idealized, impression of how the structure will appear after it is constructed.

Table 3: Definitions given at ODLIS for the words ubiquitously found by LHN approach as most similar to several nonrelated words.

On the contrary, all words found by the CCS approach are related to the target ones, and indeed in all cases the target word appears in the definition of the word found. For instance, **original**: the word “original” appear in ODLIS as: “In literature, a **work** as written by the author or in the author’s own words. In art, a finished **work** as completed by the artist and ready for reproduction, the phrase “direct edition” appears as: An edition of a **work** for which the author provides the publisher with camera-ready copy produced on a computer with the aid of word processing software. Used mainly for **works** that cannot be produced economically from type.

5 Cosine-distance closeness centrality

Here I propose an analogous of the closeness centrality index based on the CCD between a pair of nodes in a graph.

Definition 14 Let D_{vw} be the CCD between the nodes v and w . Then,

$$C_v = \left(\sum_{w=1}^n D_{vw} \right)^{-1}, \quad (5.1)$$

is the communicability closeness centrality (CCC) of the vertex v .

The closeness centrality measures, such as CC and RCC, are designed to account for the importance of a vertex in a graph in terms of its proximity to the rest of the vertices. Then, these indices are useful in comparing different vertices in the same

graph. However, I will analyze the change of the centrality of a given vertex in a graph when the size of the graph growth. The main goal of this exercise is to show some significant differences between CCC and the “classical” closeness centrality measures. I will start by the analysis of cycle graphs.

Lemma 3 *Let C_n be a cycle graph with n nodes. Let $I_\nu(z)$ be the modified Bessel function of the first kind, and let*

$$\tilde{C}_v(C_n) = \frac{I_0(2)}{2(nI_0(2) - e^2)}. \quad (5.2)$$

Then,

$$\lim_{n \rightarrow \infty} \frac{C_v(C_n)}{\tilde{C}_v(C_n)} = 1. \quad (5.3)$$

Proof The communicability between a pair of nodes v and w and the subgraph centrality of a given node in C_n are, respectively

$$G_{vw}(C_n) = \frac{1}{n} \sum_{j=0}^{n/2} \exp\left(2 \cos\left(\frac{2j\pi}{n}\right)\right) \cos\left(\frac{2j\pi(v-w)}{n}\right), \quad (5.4)$$

$$G_{vv}(C_n) = \frac{1}{n} \sum_{j=0}^{n/2} \cos\left(\frac{2j\pi}{n}\right). \quad (5.5)$$

Let

$$\tilde{G}_{vw}(C_n) := \frac{1}{\pi} \int_0^\pi \exp(2 \cos(\phi)) \cos(\phi(v-w)) d\phi = I_{d_{v,w}}(2), \quad (5.6)$$

$$\tilde{G}_{vv}(C_n) := \frac{1}{\pi} \int_0^\pi 2 \cos(\phi) d\phi = I_0(2), \quad (5.7)$$

where $d_{v,w}$ is the shortest path distance between the two nodes and $\phi = 2\pi j/n$. Then, we define

$$\cos \tilde{\theta}_{vw}(C_n) := \frac{\tilde{G}_{vw}(C_n)}{\sqrt{\tilde{G}_{vv}(C_n) \tilde{G}_{ww}(C_n)}} = \frac{I_{d_{v,w}}(2)}{I_0(2)}. \quad (5.8)$$

It can be easy to see that $\lim_{n \rightarrow \infty} \tilde{G}_{vw}(C_n)/G_{vw}(C_n) = 1$ and $\lim_{n \rightarrow \infty} \tilde{G}_{vv}(C_n)/G_{vv}(C_n) = 1$, such that $\lim_{n \rightarrow \infty} \cos \tilde{\theta}_{vw}(C_n)/\cos \theta_{vw}(C_n) = 1$.

Therefore, let

$$\tilde{C}_v(C_n) = \frac{1}{\sum_{w \neq v} 2 - 2 \cos \tilde{\theta}_{vw}(C_n)} = \frac{1}{2(n-1) - 2 \sum_{w \neq v} \cos \tilde{\theta}_{vw}(C_n)}. \quad (5.9)$$

Let us consider again that the cycle if of even length, such that

$$\sum_{w \neq v} \cos \tilde{\theta}_{vw}(C_n) = \frac{2}{I_0(2)} \sum_{k=1}^{n/2} I_k(2). \quad (5.10)$$

Then, because $e^z = I_0(z) + 2 \sum_{r=1}^{\infty} I_r(z)$, we have that

$$\lim_{n \rightarrow \infty} \tilde{C}_v(C_n) = \frac{I_0(2)}{2(nI_0(2) - e^2)}, \quad (5.11)$$

which finally proves the result. \square

Example 1 Let us consider the cycles C_n for $4 \leq n \leq 12$ and obtain the value of $\tilde{C}_v(C_n)$ using the previous result as well as $C_v(C_n)$ using the function `expm(A)` implemented in Matlab. The results are given in Table 4. First, it is observed that the values of $\tilde{C}_v(C_n)$ converge to those of $C_v(C_n)$ when the sized of the graph is still relatively small.

In this case the three closeness centralities, CC, RCC and CCC, drop with the increment of the graph size. The fastest drawing is observed for CC where $CC_v(C_n) = \frac{4}{n^2}$ if n is even, or $CC_v(C_n) = \frac{4}{n^2 - 1}$ if it is odd. It is followed by RCC where $RCC_v(C_n) = \frac{6}{n^2 - 1}$ and finally by CCC where $\tilde{C}_v(C_n) \approx \frac{2.2796}{4.5592n - 14.7781}$.

n	$C_v(C_n)$	$\tilde{C}_v(C_n)$	$CC_v(C_n)$	$RCC_v(C_n)$
4	0.5575465980	0.6591120482	1/4	0.4000
5	0.2799080062	0.2843176590	1/6	0.2500
6	0.1809535106	0.1812515986	1/9	0.1714
7	0.1330057818	0.1330283840	1/12	0.1250
8	0.1050712617	0.1050730010	1/16	0.0952
9	0.0868265836	0.0868267141	1/25	0.0750
10	0.0739798487	0.0739798581	1/30	0.0606
11	0.06444446458	0.06444446465	1/36	0.0500
12	0.0570867727	0.0570867727	1/42	0.0419

Table 4: Values of the cosine distance closeness centrality obtained using Matlab function “`expm(A)`”, $C_v(C_n)$, as well as using the results of Lemma 3, $\tilde{C}_v(C_n)$. The values of the CC and RCC are also shown for comparison.

Let us now consider the complete graphs with n vertices.

Lemma 4 *Let K_n be the complete graph with n nodes. Then,*

$$C_v(K_n) = \frac{e^n + n - 1}{2n(n - 1)}. \quad (5.12)$$

Proof The eigenvalues of the adjacency matrix of K_n are $n - 1$ with multiplicity 1 and -1 with multiplicity $n - 1$. We thereby have

$$G_{vv} = \frac{1}{ne} (e^n + n - 1), \quad (5.13)$$

$$G_{vw} = \frac{1}{ne} (e^n - 1), \quad (5.14)$$

therefore, the cosine of the communicability angle between any pair of vertices is

$$\cos \theta_{vw} = \frac{e^n - 1}{e^n + n - 1}. \quad (5.15)$$

Then the result follows by substitution in the formula for the CCC. \square

Example 2 Let us now consider all the complete graphs with $3 \leq n \leq 12$. In Fig. 5.1 we illustrate the values of CCC for a vertex of K_n as well as the values of CC and RCC. The classical closeness centrality takes values $CC_v(K_n) = (n-1)^{-1}$ which clearly decays to zero when $n \rightarrow \infty$. The shortest path distance between two nodes in K_n is always one. In the case of the RCC there are two competing factors. When the size of the graph increases the number of vertices that can be reached in one step from a given node also increases. However, due to the increase in the number of walks, the length of these edges drops as $\Omega_{ij}(K_n) = 2/n$. The resulting effect is a domination of the first of the two mentioned factors, which means that RCC drops with the increase of size, $RCC_v(K_n) = n/(2(n-1))$, reaching the asymptotic value of $1/2$ as the size growth. As a result of the second factor, RCC drops more slowly than CC as can be seen in Fig. 5.1. Finally, in the case of CCC the competition between the two previously mentioned factors is won by the drop of the edge length, which decays exponentially in this case, $D_{uv} = \frac{2n}{e^n + n - 1}$. As a result, CCC increases monotonically with the size of the complete graphs. From an application point of view it means that a node in K_n becomes more central as n growth because it can reach more neighbors with relatively little effort due to the contraction of the edge length separating them. Such contraction of edge lengths is due to the fact that more walks exists to go from one vertex to another, which can be used as ways of communication between such pair of vertices. Geometrically this means that the radius of the hypersphere in which K_n is embedded decays exponentially as n increases. Empirically we have found that $R \approx 2.167 \exp(-0.3856n)$ with a correlation coefficient of 0.9981 for $4 \leq n \leq 12$.

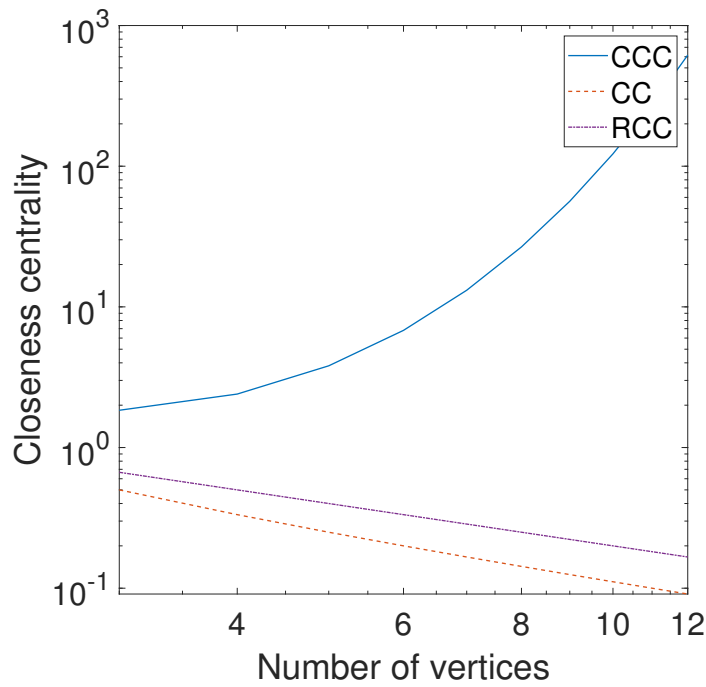


Fig. 5.1

Finally, I will illustrate the behavior of CCC on star graphs of different sizes.

Lemma 5 *Let $S_{1,n-1}$ be the star graph with n nodes and central node labeled by 1. Then,*

$$C_1(K_{1,n-1}) = \frac{\Upsilon^{1/2} \Sigma^{1/2}}{(n-1)(2\Sigma^{1/2}\Upsilon^{1/2} - 2\Phi)}, \quad (5.16)$$

$$C_j(K_{1,n-1}) = \frac{\Upsilon^{1/2} \Sigma^{1/2}}{2n^2 \Sigma^{1/2} - 4n \Sigma^{1/2} + 2\Sigma^{3/2} - \Phi \Upsilon^{1/2}}, \quad (5.17)$$

where

$$\Upsilon = \cosh(\sqrt{n-1}) + n - 2, \quad (5.18)$$

$$\Phi = \sinh(\sqrt{n-1}), \quad (5.19)$$

$$\Sigma = \cosh(\sqrt{n-1}). \quad (5.20)$$

Proof The communicability between the different pairs of nodes in $K_{1,n-1}$ are

$$G_{1w}(K_{1,n-1}) = \frac{1}{\sqrt{n-1}} \sinh(\sqrt{n-1}), \quad (5.21)$$

$$G_{vw}(K_{1,n-1}) = \frac{1}{n-1} (\cosh(\sqrt{n-1}) - 1), \quad (5.22)$$

and the subgraph centrality of these vertices are

$$G_{11}(K_{1,n-1}) = \cosh(\sqrt{n-1}), \quad (5.23)$$

$$G_{ww}(K_{1,n-1}) = \frac{1}{n-1} (\cosh(\sqrt{n-1}) + n - 2). \quad (5.24)$$

Therefore, we have

$$\cos \theta_{1w} = \frac{\sinh(\sqrt{n-1})}{\sqrt{\cosh(\sqrt{n-1})} \sqrt{\cosh(\sqrt{n-1}) + n - 2}} = \frac{\Phi}{\Sigma^{1/2} \Upsilon^{1/2}}, \quad (5.25)$$

$$\cos \theta_{vw} = \frac{\cosh(\sqrt{n-1}) - 1}{\cosh(\sqrt{n-1}) + n - 2} = \frac{\Sigma - 1}{\Upsilon}, \quad (5.26)$$

from which the results follow by substitution into the definition of CCC. \square

Example 3 Here I consider star graphs with $4 \leq n \leq 100$ for which I calculate $C_1(K_{1,n-1})$ and $C_2(K_{1,n-1})$. The comparison with the analogous of CC (notice that RCC is identical to CC because the graph is a tree) is straightforward because $CC_1(K_{1,n/1}) = (n-1)^{-1}$, and $CC_j(K_{1,n/1}) = (2n-3)^{-1}$. That is, for CC the centrality of both nodes decays with the number of vertices. Here the edge length is always equal to one, such that the only effect observed is the fact that a walker at a vertex in a star encounters more vertices to visit as n increases. However, when we consider CCC we find again two competing factors: (i) the increase in the number of first and/or second neighbors, and (ii) the change of the edge length due to the increase in the number of walks connecting pairs of vertices. The resulting effect of both factor is nonmonotonic on the number of vertices as can be seen in Fig. 5.2(a), where CCC first drop with the number of vertices and then growth for $n \geq 23$. The difference in the CCC of the central and pendant vertices also increases with the growth of the graph size. The reason for the nonmonotonic behavior observed resides in the fact that here the communicability distance between pairs of vertices decays still exponentially, but not so fast as for the case of complete graphs. Without loss of generality let me focus on the central node labeled as 1. Instead of using the more difficult to interpret expression for $\cos \theta_{1w}$ that we have found before, let us proceed as follow. We can obtain an empirical relation between between the communicability distance and n by finding the best nonlinear regression model. It results to be of the form: $D_{1,j} \approx ae^{bn} + ce^{dn}$, where I have found that $a \approx 0.8896$, $b \approx -0.0472$, $c \approx -0.5259$, and $d \approx -0.2606$. Similar results are also obtained for $D_{i,j}$ when $i \neq j \neq 1$. The result of the “exact” values of these distances and that of the best fitting are given in Fig. 5.2(b). Thus, what happens in star graphs is that for relatively small number of vertices the contraction in the communicability distance is not enough to overcome the effect of the increase in the number of paths of length one and two. That is, the centrality of the vertex 1, for instance, decays when n increases in this region because the vertex has to make a greater “effort” to “contact” a larger number of nearest neighbors. However, after certain size, due to the contraction of the edge lengths between this vertex and its neighbors, the number of nearest neighbors that can be visited from this vertex is bigger and bigger without increasing too much the “effort” that it takes to visit them.

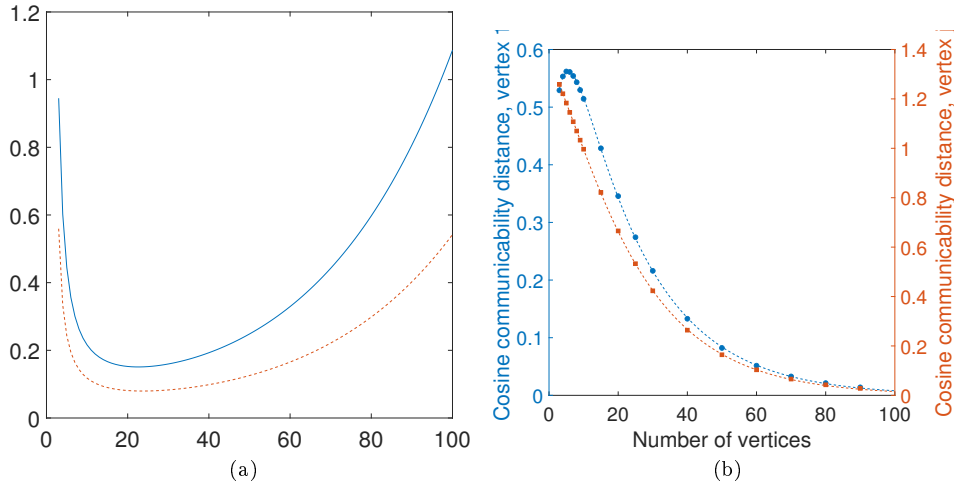


Fig. 5.2: (a) Plot of the communicability closeness centrality for the central node (continuous line) and the pendant node (broken line) of star graphs with n vertices. (b) Plot of the cosine communicability distance from the vertex 1 to a nearest neighbor (left) and between a pair of adjacent vertices $i \neq j \neq 1$ (right).

In closing, we have seen that CCC has very different behavior for different types of graphs and that they differ significantly from those observed for CC and RCC.

6 Discriminant power of CCC

Bao and Zhang [13] have recently analyzed the discriminant power of several centrality measures, including the CC one. They quantify the number of pairs of vertices that are differentiated by a given centrality measure relative to the total number of pairs as the discriminant power of the centrality. Here I analyze the discriminant power of the CCC and the analogous closeness centrality CC and RCC, based on their capacity to discriminate the nonidentical vertices in a graph. Let me start with the following.

Example 4 Let us consider the labeled graph illustrated in Fig. 6.1. As the graph has 7 vertices it has 21 pairs of vertices. Let us check for every of these 21 pairs of vertices if there is a permutation matrix P that interchange these two vertices, such that: $PAP^T = A$. We have found that there is only one of such permutations vertices, which interchanges the vertices labeled as 3 and 6, which are colored in black in Fig. 6.1. Therefore, there is at least one nontrivial automorphism of the graph that swap these two vertices. These vertices are name “similar” in graph theory, but here I propose to call them “identical” in order to avoid confusion with the quantitative concept of vertex similarity used before.

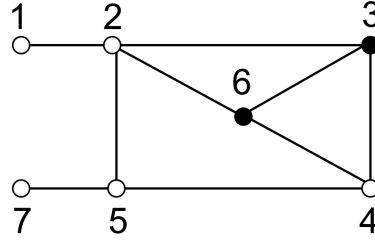


Fig. 6.1: Illustration of a labeled graph used in the Example.

Let us now consider the values of three closeness centrality measures studied here for the vertices in the graph represented in Fig. 6.1. The classical CC does not differentiate among the vertices labeled by numbers 3, 4 and 6 (see Table 5). That is, there are three pairs of indistinguishable vertices according to this centrality. The resistance closeness centrality identifies the pair 3, 6 as having the same centrality but it also identifies another pair of vertices as indistinguishable, which corresponds to the pair 4, 5. As we have seen these vertices are not identical because there is not a permutation matrix that interchange them leaving unchanged the adjacency matrix. As can be seen in the third column of see Table 5 the CCC only identifies one pair of vertices as indistinguishable and they coincide with the identical vertices previously found.

No.	CC_v	RC_v	C_v
1	0.0769	0.0991	0.1397
2	0.1250	0.1967	0.2789
3	0.1000	0.1702	0.2395
4	0.1000	0.1739	0.2375
5	0.1111	0.1739	0.2311
6	0.1000	0.1702	0.2395
7	0.0714	0.0930	0.1244

Table 5: Values of the communicability closeness centrality (CCC) as well as of the standard closeness centrality and of the resistance closeness centrality for the nodes in the graph illustrated in Fig. 6.1.

Let us first designate $\hat{x}_j = \frac{x_j}{\|x_j\|}$, such that $D_{vw} = \|\hat{x}_v - \hat{x}_w\|^2$ and let us call communicability cosine fairness to

$$F_v = \sum_{w=1}^n D_{vw}. \quad (6.1)$$

Then, we have the following.

Lemma 6 *Let v and w be two different vertices of a graph G . Then, if $F_v(G) = F_w(G)$,*

$$(\hat{x}_v - \hat{x}_w) \cdot (\hat{x}_1 + \cdots + \hat{x}_n) = 0, \quad (6.2)$$

meaning that either $\hat{x}_v = \hat{x}_w$ or that $\hat{x}_v \cdot R = \hat{x}_w \cdot R$, where $R = \hat{x}_1 + \dots + \hat{x}_n$ and $a \cdot b$ indicates inner product.

Proof We can write

$$\begin{aligned} F_v(G) &= F_w(G) \\ \|\hat{x}_v - \hat{x}_1\|^2 + \dots + \|\hat{x}_v - \hat{x}_n\|^2 &= \|\hat{x}_w - \hat{x}_1\|^2 + \dots + \|\hat{x}_w - \hat{x}_n\|^2 \\ \hat{x}_v \cdot \hat{x}_1 + \hat{x}_v \cdot \hat{x}_2 + \dots + \hat{x}_v \cdot \hat{x}_n &= \hat{x}_w \cdot \hat{x}_1 + \hat{x}_w \cdot \hat{x}_2 + \dots + \hat{x}_w \cdot \hat{x}_n. \end{aligned} \quad (6.3)$$

Because the vectors \hat{x}_v and \hat{x}_w are respectively in the LHS and RHS of the sums and recalling that $\hat{x}_v \cdot \hat{x}_v = 1$ we have $F_v(G) - F_w(G) = 0$ implies that

$$(\hat{x}_v - \hat{x}_w) \cdot (\hat{x}_1 + \dots + \hat{x}_n) = 0. \quad (6.4)$$

Because $0 \leq M_{vw} \leq 1$, the angle between every pair of position vectors is not larger than 90° , which implies that $\hat{x}_1 + \dots + \hat{x}_n > 0$, which finally proves the result. \square

Remark 2 If CCC discriminate vertices up to automorphism it should be the case that $C_v = C_w$ implies that $AP = PA$ for a permutation matrix P that swap the vertices v and w . This last condition is fulfilled if $\varphi_v = \varphi_w$. The fact that $\hat{x}_v \cdot R = \hat{x}_w \cdot R$, does not imply mathematically that $\hat{x}_v = \hat{x}_w$ and consequently does not imply that $\varphi_v = \varphi_w$. Therefore, this condition does not imply that if $F_v(G) = F_w(G)$ there is a permutation that swap the two vertices v and w while preserving the adjacency matrix. Let us then consider the second case, that is when $\hat{x}_v = \hat{x}_w$. First, let us consider that $x_v = x_w$ which means that $e^{A/2}(\varphi_v - \varphi_w) = 0$. Therefore, it implies that $\varphi_v = \varphi_w$. Also because $G_{vv} = x_v \cdot x_v$, the fact that $x_v = x_w$ also implies that $G_{vv} = G_{ww}$. In closing, if $\hat{x}_v = \hat{x}_w$ we have that

$$e^{A/2} \left(\frac{\varphi_v}{G_{vv}} - \frac{\varphi_w}{G_{ww}} \right) = 0, \quad (6.5)$$

which necessarily implies that $\varphi_v = \varphi_w$.

Because it is not always necessarily true that $C_v = C_w$ implies that the two vertices are identical we then studied all 11,117 connected graphs with 8 vertices and identified the number of graphs with a given number of pairs of identical vertices. For instance, there are 3552 graphs with 8 vertices which have no pair of identical vertices (see next section). There are 2825 graphs with only one pair of identical vertices, 1913 with two pairs, and so on. In Table 6 I give the number of graphs with a given number of distinguishable pairs of vertices for the three closeness centralities studied here. For instance, there are only 12 graphs for which CC give different values for all their vertices, this number increases up to 2823 for RCC, but only CCC identifies all graphs having no pair of identical vertices. As can be seen in Table 6 the CCC identifies all graphs with any number of pairs of identical vertices among the graphs with 8 vertices. This performance is not observed even for other indices based on the exponential of the adjacency matrix, such as the subgraph centrality (SC), for which I shown the number of identical pairs of vertices identified by this centrality in Table 6. Both CC and RCC are far from discriminate all the pairs of nonidentical vertices in these graphs.

#pairs	CC	RCC	CCC		SC	real
0	12	2823	3552		3494	3552
1	72	2042	2825		2813	2825
2	194	1387	1913		1888	1913
3	279	1206	1449		1440	1449
4	207	489	586		586	586
5	79	148	167		167	167
6	41	75	91		91	91
7	123	201	218		216	218
8	129	180	185		185	185
9	15	19	20		20	20
10	7	8	8		8	8
11	12	14	14		14	14
12	38	39	39		39	39
13	4	5	5		5	5
15	7	7	7		7	7
16	25	25	25		25	25
21	3	3	3		3	3
28	10	10	10		10	10

Table 6: Number of connected graphs with 8 vertices which are identified by the three closeness centrality studied here, plus the subgraph centrality, as having a given number of equivalent vertices. The last column (real) indicates the number of graphs having a given number of identical vertices. For instance, there are 3552 graphs with 8 vertices that have 0 pair of identical vertices, i.e., they are identity graphs. From those, CC identifies 12, RCC 2823, SC 3494, and CCC identifies all of them.

6.1 Graphs with no pair of identical vertices

Let me recall that a graph is called asymmetric or identity if it does not contain any pair of identical vertices [27]. That is, if there is no nontrivial automorphism for the graph. Asymmetric graphs are of relevance for the study of graph controllability [51–54]. For instance, it was proved that the class of essentially controllable graphs form a strict subset of asymmetric graphs [51]. Additionally, identity graphs have been proposed as “the mathematical structure of the World” in the “World as a graph” proposal of Dipert [55] (see also [56]). The idea of “World as a graph” is that if the World is a graph, then it has to be asymmetric to avoid that two spatio-temporal points can be identical and therefore can be swapped.

There are no asymmetric graphs with $2 \leq n < 6$, but there are 8 asymmetric graphs with 6 vertices, which are illustrated in Fig. 6.2).

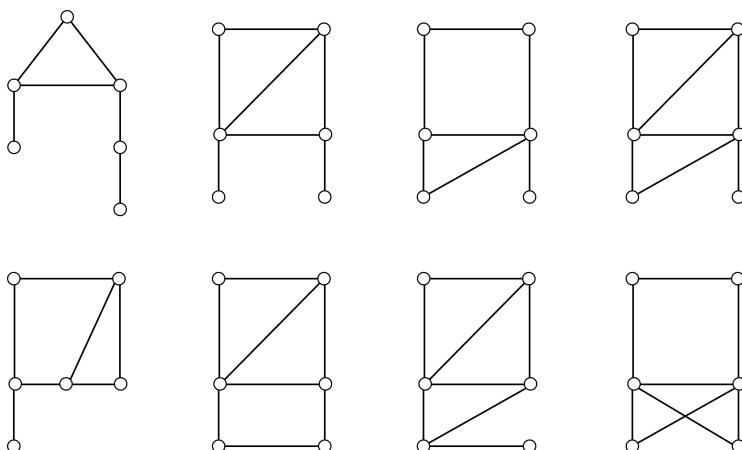


Fig. 6.2: Illustration of the asymmetric (identity) eight graphs with 6 vertices, i.e, those graphs with 6 vertices for which $\text{Aut}(G) = 1$.

I calculated the CCC for all the vertices in these 8 asymmetric graphs with $n = 6$ and found that this centrality distinguishes all their vertices. That is, let \mathcal{A}_n be the set of all asymmetric graphs with n vertices, we have observed that if $v \neq w$ then $C_v \neq C_w$ for all $v, w \in \mathcal{A}_6$. This is not the case for the standard CC, which does not distinguish all the vertices in any of these asymmetric graphs, nor for the RCC, which only finds all vertices different for 4 out of the 8 asymmetric graphs. Although not a closeness centrality we also compared the results with the SC, which distinguishes all vertices in these 8 asymmetric graphs with 6 vertices.

I now extend these calculations to all asymmetric graphs with $7 \leq n \leq 9$ and the results are displayed in Table 7. It is known that the cardinality of the sets \mathcal{A}_n are 8, 144, 3552, 131452 for $n = 6, 7, 8, 9$, respectively (see integer sequence A124059 in the Encyclopedia of Integer Sequences [57]). As can be seen in Table 7 the CCC differentiates all the vertices in 100% of the asymmetric graphs with $6 \leq n \leq 9$. That is, if $v \neq w$ then $C_v \neq C_w$ for all $v, w \in \mathcal{A}_{n \leq 9}$. The standard CC differentiates very poorly the vertices of these asymmetric graphs, distinguishing only 2.1% of the asymmetric graphs with 7 vertices, 0.36% of those with 8 vertices and 0.09% of those with 9 vertices. The situation improves when we consider the resistance closeness centrality which differentiates 93.75%, 79.53% and 96.33% of asymmetric graphs with 7, 8 and 9 vertices, respectively. Here again, the SC is the second best, after CCC, and differentiates 100%, 97.83% and 98.3% of asymmetric graphs with 7, 8 and 9 vertices, respectively.

	$n = 6$	$n = 7$	$n = 8$	$n = 9$
CC	0	3	13	116
RCC	4	135	2824	126,626
CCC	8	144	3552	131,452
$\#\mathcal{A}_n$	8	144	3552	131,452

Table 7: Number of asymmetric (identity) graphs whose vertices are differentiated by the different centrality measures for $6 \leq n \leq 9$.

A particularly interesting class of graphs is formed by those having at least one pair of pseudosimilar vertices [25, 26] (hereafter I will call them “pseudoidentical”). That is, these are graphs for which a pair of vertices v, w exists such that $G - v$ is isomorphic to $G - w$ but there is no automorphism transforming v onto w . From the perspective of the current work, these pairs of vertices could be seen as difficult ones to be distinguished by any centrality measure. Two examples of asymmetric graphs with a pair of pseudoidentical vertices are the graphs G and H illustrated in Fig. 6.3. Both graphs have no automorphism but the identity one. In graph G the vertices 4 and 6 are pseudoidentical as well as the vertices 5 and 7 in H .

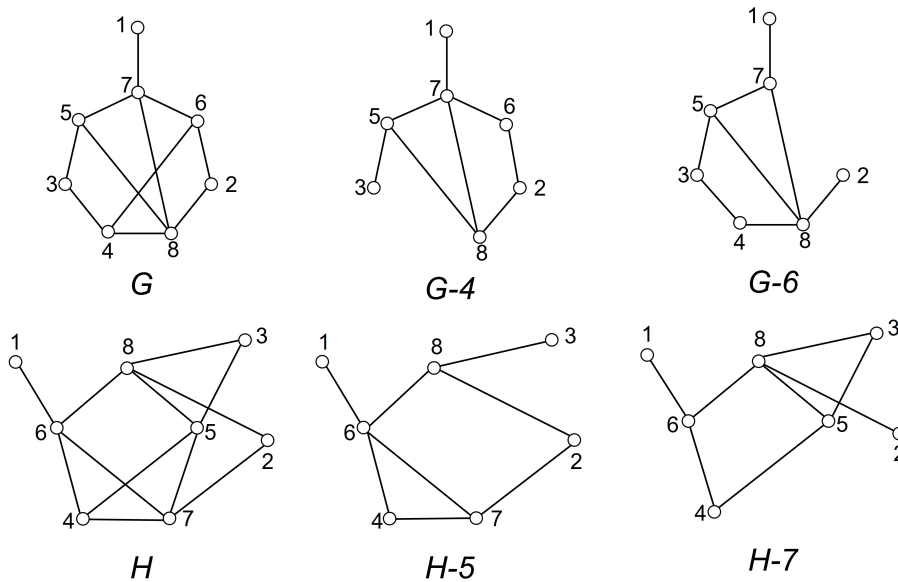


Fig. 6.3: Illustration of two pseudoidentity graphs. In graph G the vertices 4 and 6 are those that if removed the resulting graphs $G - 4$ and $G - 6$ are isomorphic (see first line of the figure). For the graph H such vertices are 5 and 7 and the resulting isomorphic graphs are $H - 5$ and $H - 7$.

In Table 8 we give the values of CCC , RCC and SC for every vertex in the two graphs illustrated in Fig. 6.3. I do not display the results for the classical CC because

it is highly degenerated in most of the cases, i.e., many pairs of nonidentical vertices have the same values of the CC. Although the RCC fails to differentiate all the vertices of these two graphs, the pairs of vertices for which RCC gives the same values do not coincide with the pseudoidentical ones. That is, in graph G the pseudoidentical vertices are labeled 4 and 6 while the RCC produces identical values for vertices 5 and 6. The same happens for graph H where the RCC is degenerated for vertices 6 and 8 while the pseudoidentical ones are 5 and 7. This situation is repeated across the asymmetric graphs with 8 vertices where we have identified 36 graphs having one pair of pseudoidentical vertices. The vertices of 22 of these 36 graphs are distinguished by the RCC and in the remaining cases the pairs with similar values of this centrality do not necessarily coincide with the pseudoidentical ones. A different situation occurs with the SC which have the same value for the pairs of pseudoidentical vertices in both graphs illustrated in Fig. 6.3 (see table 8). However, the SC does not only produce degeneration for these pairs of pseudoidentical vertices but occasionally also in other pairs, such as pair 6,8 in graph H . This happened in 14 out of the 36 asymmetric graphs with one pair of pseudoidentical vertices.

No.	graph G			graph H		
	C_v	RC_v	SC_v	C_v	RC_v	SC_v
1	0.1151	0.0888	1.7342	0.1172	0.0904	1.7459
2	0.1599	0.1367	2.7588	0.1831	0.1445	2.9435
3	0.1460	0.1338	2.6019	0.1687	0.1391	3.1937
4	0.1944	0.1675	3.7434	0.2344	0.1766	5.0812
5	0.2014	0.1692	4.1330	0.2631	0.1952	6.2944
6	0.1952	0.1692	3.7434	0.2604	0.1978	5.6427
7	0.2303	0.1902	5.1577	0.2654	0.1987	6.2944
8	0.2449	0.1931	5.4313	0.2567	0.1978	5.6427

Table 8: Values of centrality measures for the two pseudoidentity graphs illustrated in Fig. 6.3. The values of the CCC and RCC are provided together with those of the subgraph centrality (SC). Identical values are marked in bold (first pair) and gray (second pair). Notice that no pair of identical vertices exist in these two graphs.

I calculated the difference between the CCC of every pair of vertices in the two graphs illustrated in Fig. 6.3. In both cases it is found that the smallest difference is observed for the pair of pseudoidentical vertices. Unfortunately, this is not always true, as for the 36 asymmetric graphs with a pair of pseudoidentical vertices, the smallest difference between the CCC of every pair of vertices coincides with the pseudoidentical vertices in 27 occasions, but it does not coincide in 9.

Let me now consider some other examples of known identity graphs which are illustrated in Fig. 6.4, which have from 12 to 222 vertices.

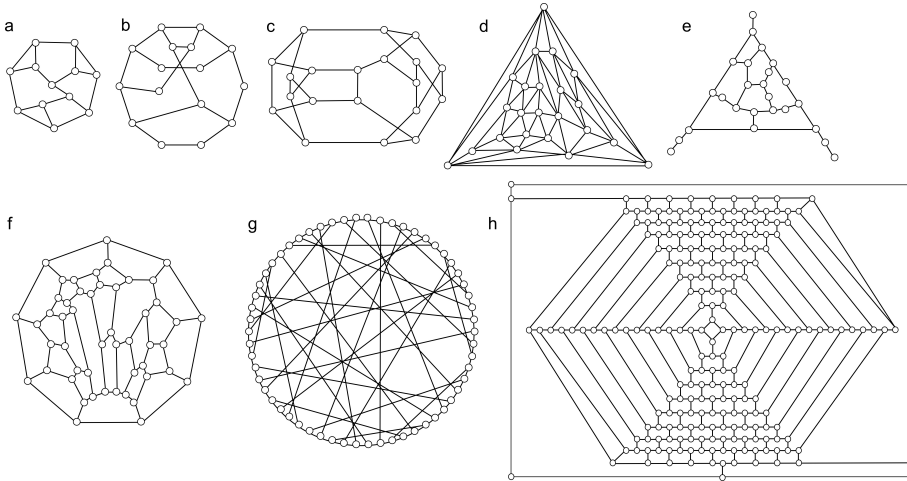


Fig. 6.4: Illustration of the structure of eight “classical” examples of identity graphs, i.e., those having no symmetry but the identity one.

In Table 9 I give the number of pairs of vertices which CC, RCC and CCC identify as equivalent. As can be seen CCC differentiates all the vertices in the eight graphs. RCC identifies as equivalent a pair of vertices in the Frucht graph but in general distinguishes very well the vertices in these graphs. Finally, CC only gives different values for the vertices of the (10,3)-incidence graph illustrated.

	name	n	CC	RC	CCC
a	Frucht graph	12	13	1	0
b	(10,3)-incidence graph	16	0	0	0
c	20-snark graph	20	36	0	0
d	Kittell graph	23	18	0	0
e	Walther graph	25	5	0	0
f	cubic nonhamiltonian graph	50	21	0	0
g	(3,9)-cage graph	58	498	0	0
h	Gardner graph	222	59	0	0

Table 9: Number of pairs of vertices with the same value of the closeness centrality measures studied here for eight “classical” examples of identity graphs. For instance, for the Frucht graph CC identifies 13 pairs of equivalent vertices and RCC identifies one pair, while this graph has zero pairs of identical vertices, which is correctly identified by CCC.

Note 1 A note of caution should be raised here about numerical accuracy. In some cases, the values of the centrality can be very close for different pairs of vertices and the precision of the calculations should be increased. However, it also must be taken into account what is the accuracy of the numerical methods used to compute the corresponding centrality. For instance, in the case of CCC in the Gardner graph we have to consider accuracy up to 15 decimal places and using $\sum_{w=1}^n D_{vw}$ instead of

its reciprocal in order to avoid problems with rounding errors. Therefore, I would like to remark that more numerical analysis is needed before giving conclusive statements about the true discriminant power of CCC.

6.2 Graphs with at least one pair of identical vertices

In Fig. 6.5 I illustrate eight graphs in which there are one (a), two (b), three (c), four (d), five (e), six (f), seven (g) and eight (h) pairs of identical vertices, respectively. The number of identical pairs of vertices was identified computationally by obtaining all possible permutations of the 8 vertices and checking which ones obeys that $PAP^T = A$. The groups of identical nodes are identified in the Figure with different colors. For instance, in graph b) there are two pairs of identical vertices, 1,2 colored in black and 6,7 in gray. For larger number of pairs it is possible that such pairs are grouped in clusters. For instance, in c) where there are three pairs, they are formed by the vertices 3,4,5 (3,4; 3,5; 4,5). It could be in other cases that the pairs are isolated and not grouped in this way.

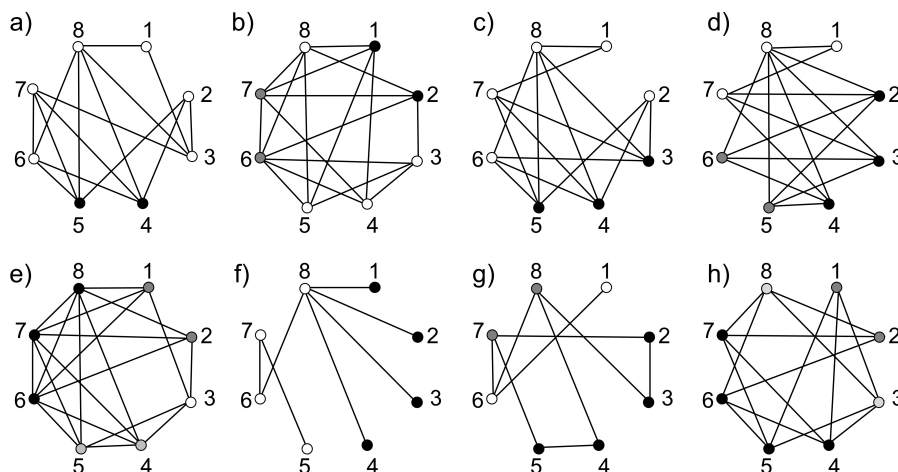


Fig. 6.5: Illustration of graphs with 8 vertices having one (a), two (b), three (c), four (d), five (e), six (f), seven (g) and eight (h) pairs of identical vertices, respectively. Every group of identical vertices is colored the same. For instance, in h) there are found vertices in black, which form 6 pairs of identical vertices, plus one pair in dark gray and another pair in pale gray, summing 8 pairs of identical vertices.

In Table 10 I give the values of CCC for every vertex in these eight graphs. The pairs of nodes for which $C_v = C_w$ coincide in all cases with the identical nodes existing in these graph and no other pair of nodes have the same values of CCC. This situation is very different for the other closeness centrality measures. For instance, for graph a) CC gives the same value $CC_j = 0.1, j = 3, \dots, 7$ for five vertices, such that there are 10 pairs of vertices with the same values of CC. The RCC identifies, apart from the

pair 4,5 for which $RC_4 = RC_5 \approx 0.2689$, also the pair 6,7 as having the same value of this centrality, $RC_6 = RC_7 \approx 0.2647$. The SC identifies the vertices 4,5,7 with the same value, $SC_4 = SC_5 = SC_7 \approx 8.0135$. In the case of the graph b) the CC identifies 11 pairs of equivalent vertices (vertices 1-5 and 6,7), while the RCC identifies three pairs (1,2; 3,4 and 6,7) as well as SC which identifies 1,2; 3,5 and 6,7. In graph c) CC and RCC identify the same 10 pairs of equivalent vertices (3-7), they also coincide in identifying 15 pairs (2-7) for graph d) and 7 pairs (1-3; 4,5 and 6-8) for graph e). The graph f) is a tree, which is shown to illustrate that even in very simple graphs CC identifies different set of vertices than CCC. In this case CC (RCC is identical to CC for trees) identifies vertex 7 as equivalent to vertices 1-4, which are the identical ones. In graphs g) and h) CC identifies 9 and 16 equivalent nodes, respectively, while RCC identifies correctly only the identical vertices in g) and h). However, in g) SC identifies 9 vertices which coincide with those identified by CC. If we consider the sum of the rows of the LHN matrix \mathcal{N} (Eq. 4.1) as a centrality measure, then, the graph h) is an example where it fails to identify only the identical vertices because it identifies 16 pairs of equivalent vertices (1,2 and 3-8) instead of 8. More examples exist among the graphs with 8 vertices.

vertex	a	b	c	d	e	f	g	h
1	0.2146	0.7556*	0.2409	0.2858	1.3450*	0.1121*	0.0989	0.2664*
2	0.3361	0.7556*	0.3512	0.6232*	1.3450*	0.1121*	0.1211*	0.2664*
3	0.3909	0.6811	0.4993*	0.6232*	1.0402	0.1121*	0.1211*	0.3738†
4	0.4370*	0.7788	0.4993*	0.6232*	1.7942†	0.1121*	0.1211*	0.4052‡
5	0.4370*	0.6788	0.4993*	0.5955†	1.7942†	0.0872	0.1211*	0.4052‡
6	0.4193	0.9668†	0.4877	0.5955†	2.2196‡	0.1330	0.1475	0.4052‡
7	0.4360	0.9668†	0.4519	0.5081	2.2196‡	0.1032	0.1467†	0.4052‡
8	0.5264	1.2170	0.5730	0.8262	2.2196‡	0.1610	0.1467†	0.3738†

Table 10: Values of CCC for the vertices in the graphs illustrated in Fig. 6.5. *, † and ‡ are used to indicate the different groups of identical vertices in every graph.

Remark 3 A note of caution should be stressed here. It is not the case that centrality measures in general should differentiate vertices up to their automorphism. For instance, CC is designed to identify those vertices in a graph which are closer to the rest of the vertices. Therefore, two vertices with the same proximity to the rest of the vertices of the graph have to have the same CC even if they are not identical.

6.3 CCC in networks

Here I consider 14 networks representing complex systems in a variety of scenarios. They include a flip-flop electronic circuit, the neuronal network of the worm *C. elegans*, the networks of connections between regions of cat and macaque cortex, a network connecting pairs of human brain that coactivate, a network indicating whether the buyers of a given political book also buy another one, the Roget thesaurus, and the ODLIS, a network representing the flights connections between USA airports, a food web in Bridge Brooks, the protein-protein interaction network of yeast, a network indicating whether two drug users have interchanged needles in a given period of time,

the network indicating whether two authors in computational geometry have published a paper together, and a network representing the Internet at the Autonomous System (more details and references can be found in the Appendix of [6]).

6.3.1 Distinguishing vertices

For all the 14 real-world networks previously described I calculated the number of pairs of identical vertices $\#\mathcal{A}$, as well as the number of vertices which are differentiated by each of the three CC studied. The first interesting result is that all neuronal systems in different species considered here are identity graphs. This is particularly interesting in the case of those neuronal systems coming from brains with bilateral symmetry, such as cat and macaque visual cortex as well as the human brain. That is, the bilateral symmetry, as well as any other kind of symmetry, is broken in these networks possibly due to functional reasons. However, the study of the implication of this network asymmetry is outside the scope of this work. As can be seen in Table 11 in all these cases RCC and CCC distinguish all pairs of vertices.

The degree of vertices redundancy in a network can be calculated as

$$\mathcal{R}(G) := \frac{2\#\mathcal{A}}{n(n-1)}, \quad (6.6)$$

which are given as percentages in Table 11. As can be seen, real-world networks have very few pairs of identical vertices in relation to their total possible number. This contrasts with the very high global symmetry observed by some of these graphs obtained from the automorphism group. For instance, $|\text{Aut}(G)| = 2.5916 \cdot 10^{24}$ for the airport network in USA and $|\text{Aut}(G)| = 1.8994 \cdot 10^{320}$ for the collaboration network in computational geometry (we have used exactly the same versions of the networks used in [18]). This means that there are much more ways of transforming the 1436 pairs of identical vertices existing in the collaboration network than symmetry operations exist to transform the 135 pairs of identical vertices in the airport network. However, as we have remarked in the Introduction with the example of K_4 and C_4 the size of the automorphism group does not indicate the number of pairs of vertices which are identical, which is what is matter in many real-world problems, such as the problem of controllability.

network	n	CC	RCC	CCC	$\#A$	$\mathcal{R}(G)$
electronic circuit	522	109	0	0	0	0
<i>C. elegans</i> neurons	280	143	0	0	0	0
cat cortex	52	43	0	0	0	0
macaque cortex	30	31	0	0	0	0
Human brain	638	469	0	0	0	0
Political books	105	42	0	0	0	0
Roget thesaurus	994	307	3	3	3	$6 \cdot 10^{-4}$
ODLIS	2898	1324	121	121	121	0.0029
USA airports	332	297	135	135	135	0.246
Food web	75	451	446	446	446	16.07
PIN yeast	2224	1451	863	863	863	0.035
Drugs	612	1060	636	636	636	0.34
Collaboration	3621	2169	1440	1436	1436	0.022
Internet_97	3015	38,816	37,893	37,893	37,893	0.85

Table 11: Number of pairs of vertices which are differentiated by the three closeness centrality studied here in the 14 real-world networks analyzed. For comparison the number of nonidentical pairs of vertices, $\#A$, in every network is also provided. For instance, $\#A = 0$ indicates that the graph is asymmetric (identity).

Let us now turn our analysis to the identical vertices in the Roget Thesaurus network. The complete enumeration of all pairs of vertices that can be swapped by a non-trivial automorphism identifies only three pairs, formed by words: “duality”-“bisection”, “celibacy”-“divorce”, and “man”-“woman”. The CC identified 307 pairs of words with equivalent closeness centrality. However, both CCC and RCC identifies three uniquely existing identical pairs of words. The pairs “duality”-“bisection” and “celibacy”-“divorce” are pairs of trivially identical words because they correspond to vertices of degree one which are connected to the same root vertex. In the first case the two words are connected to “duplication” and in the second they are connected to “marriage”. The words “man” and “woman” are connected to each other in the network and each of them is connected to other three words: “infant”, “adolescent” and “marriage”.

The second example is the network representing ODLIS which is formed by 2898 vertices and 16,376 edges, after the elimination of 5 self-loops (self-referenced words). In this case there are 121 identical pairs of words. The CCC and RCC identify correctly these 121 pairs of words, while CC identifies 1324 pairs of equivalent words. The large majority of the existing identical words correspond to trivial pairs—pairs of pendant vertices connected to the same vertex. For instance, 102 out of 121 pairs correspond to such cases, with groups of up to 8 pendant vertices connected to the same node. One of these groups is formed by the words: “book cradle”, “bookrest”, “imbrication”, “lectionary”, “polaire” and “bibliotaph”, which are all connected to the term “bibliographic item”, which have appeared previously in the analysis of words similarity in this work. There are, however, 19 pairs of words forming pairs of nontrivial identical vertices, i.e., their vertices have degree larger than one. These 19 pairs of nontrivial identical vertices are formed only by isolated pairs, that is, there are no triples, quadruples, etc. Some examples are: “chef-d’oeuvre”-“masterpiece” of degree 2; “imprimatur”-“in press” of degree 3; “americanize”-“briticize”, “elegy”-“ode”; “burlesque”-“parody”, all of degree 4; “color plate”-“monochrome plate” of degree 5; “broader term (BT or B)”-“narrower term (NT or N)” of degree 12.

6.3.2 Ranking vertices

One of the most important applications of centrality measures is the ranking of vertices in decreasing order of the given centrality. The most central vertices are then expected to play some fundamental, structural and/or dynamical, role in the network. However, in most of real-world networks, vertex centrality measures are highly correlated with the degree of the vertices, in particular for those which depends on shortest paths. For instance, if a vertex has degree k there are $k(k-1)/2$ shortest paths of length two that cross this vertex. Therefore, if k is relatively large, there are high chances that the betweenness centrality of this node is also high. Thus, we are counting duplicated information in this case. This happens, for instance, in the social network where a group of 34 members of a karate club express their friendship preferences. The nodes labeled by 1 and 34 correspond to the trainer and administrator of the club and they are visibly the most connected ones with degrees 16 and 17 respectively. A polarization is known to exist in this network as there are two factions in it: one following the administrator and the other following the trainer. In Fig. 6.6 I represent this network with vertices colored and with radii proportional to their degrees (a) and betweenness centrality (b). If we consider the CC (see Fig. 6.6 (c)) we can see that it is also biased by the degree of the nodes. The ranking according to CC is: 1, 3, 34, 32, 33, which are the vertices with degrees 16, 10, 17, 6, 12. RCC identifies 34, 1, 3, 33, and 2 as the most central vertices which have degrees: 17, 16, 10, 12, and 9. Thus, it is even more biased by degree than CC. However, CCC identifies the vertices 9, 3, 20, 32, 14 and 31 with degrees 5, 10, 3, 6, 5, and 4, but placed somehow in between the two groups in conflict in this network. This group of the most central vertices according to CCC are those for which there are a relatively large number of short walks connecting them to the rest of the vertices but a relatively small number of walks starting and ending at themselves. They are “good communicators” between the two factions existing in the network.

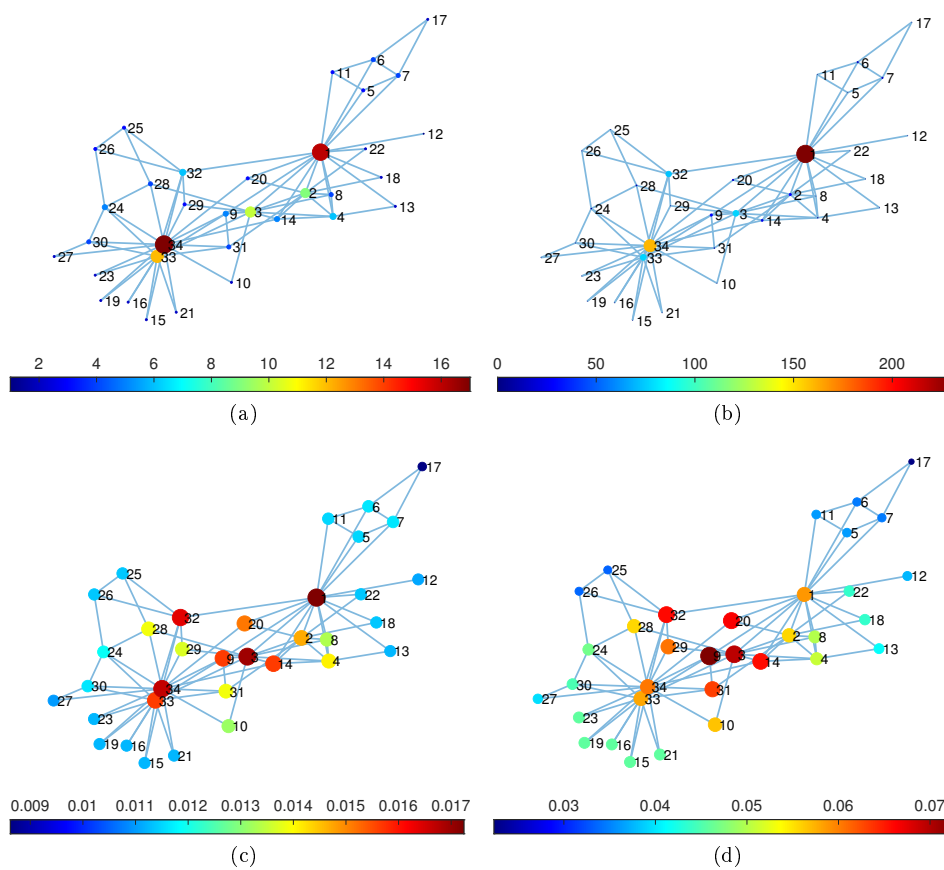


Fig. 6.6: Illustration of the degree (a), betweenness (b), closeness (c) and communicability closeness (d) centrality of the nodes in the network of a karate club, where the size and color of the vertices are proportional to the corresponding centrality.

We can now identify the most central words according to CCC in the Roget Thesaurus network, which are given in Table 12 together with the degree, CC and RCC. As can be seen CC and RCC identifies as most central the most connected words in the thesaurus, while CCC again identifies words with relatively small degree but well connected to the rest of words via relatively short walks.

degree	CC	RCC	CCC
indication (28)	inutility (25)	deterioration (27)	hope (12)
neglect (27)	neglect (27)	neglect (27)	non-preparation (13)
deterioration (27)	preparation (25)	indication (28)	success (11)
inactivity (26)	deterioration (27)	preparation (25)	danger (16)
information (25)	hindrance (24)	inutility (25)	importance (15)

Table 12: Most central words in Roget Thesaurus according to degree and the three closeness centrality studied here. The degree of the vertices is indicated in parenthesis.

These differences are even more remarked in the case of the ODLIS network as can be seen in Table 13 where CC and RCC identify the most connected vertices, while CCC identifies vertices of relatively small degree.

degree	CC	RCC	CCC
book (593)	library (566)	book (593)	nonprint (30)
library (566)	book (593)	library (566)	lacuna (8)
work (380)	publishing (368)	work (380)	obituary (10)
printing (372)	work (380)	publishing (368)	subject bibliography (13)
publishing (368)	catalog (204)	printing (372)	digital library (12)

Table 13: Most central words in ODLIS according to degree and the three closeness centrality studied here. The degree of the vertices is indicated in parenthesis.

Finally, I will focus on a network of books about US politics, where books about US politics published around the time of the 2004 presidential election and sold by the online bookseller Amazon.com are represented as vertices and edges between books represent frequent co-purchasing of books by the same buyers. The books have been classified as “liberal”, “conservative” and “neutral” as can be seen in Fig. 6.7(a). The degree of the vertices identifies three conservative books as the most central ones, followed by five liberal books and then again three conservatives and so on (see Fig. 6.7(b)). In the case of closeness measures we would expect that books placed more or less equidistant from both polarized groups, i.e., liberal and conservative, would be among the most central ones. The CC identifies the books “The Price of Loyalty”, “Rise of the Vulcans”, “The Bushes”, “Ghost Wars” and “Bush Country” as the most central ones. The books at positions 2, 3 and 4 of CC ranking are almost equidistant of both main political groups. “Rise of the Vulcans” is connected to 5 conservative books, 3 neutral and 4 liberal. “The Bushes” and “Ghost wars” both have 4, 2, 2 connections to conservative, neutral and liberal books. However, “The Price of Loyalty” and “Bush Country” are selected due to the CC bias towards high degree nodes. The first has 2, 3, 15 connections, being clearly closer to liberal books, and the second has 14, 2, 0, being clearly closer to conservative ones.

RCC is highly biased towards degree as can be seen in Fig. 6.7(c). Indeed in the top ten ranking of books by RCC we find the top most connected books in the network, and there is a high correlation between the two centralities. Indeed $RCC \sim k^{-\zeta}$ with a squared correlation coefficient of 0.97. The ranking produced by the betweenness centrality also displays bias by the degree of the vertices. It ranks at the top the books:

“The Price of Loyalty”, “The Bushes”, “Bush Country”, “Off with Their Heads” and “American Dynasty”. The books at positions 1, 4 and 5 are among the most connected ones in the network: (the first was already analyzed), “Off with Their Heads” has 22, 3, 0 connections and “American Dynasty” has 1, 2, 19 connections.

Finally we arrive at the ranking produced by CCC where the top books are: “The Bushes”, “Ghost Wars”, “Rise of the Vulcans”, “Sleeping With the Devil”, and “Why Courage Matters”, which have (4, 2, 2), (4, 1, 3), (5, 3, 4), (3, 3, 2) and (4, 1 0) connections, respectively. As can be seen in Fig. 6.7(f) all the most central books are at the central part of the network almost equidistant from both major political wings, without any bias due to the vertex degree, which is a characteristic not observed by any of the other centrality measures and illustrate one of the main characteristic features of the communicability cosine distance and derived indices.

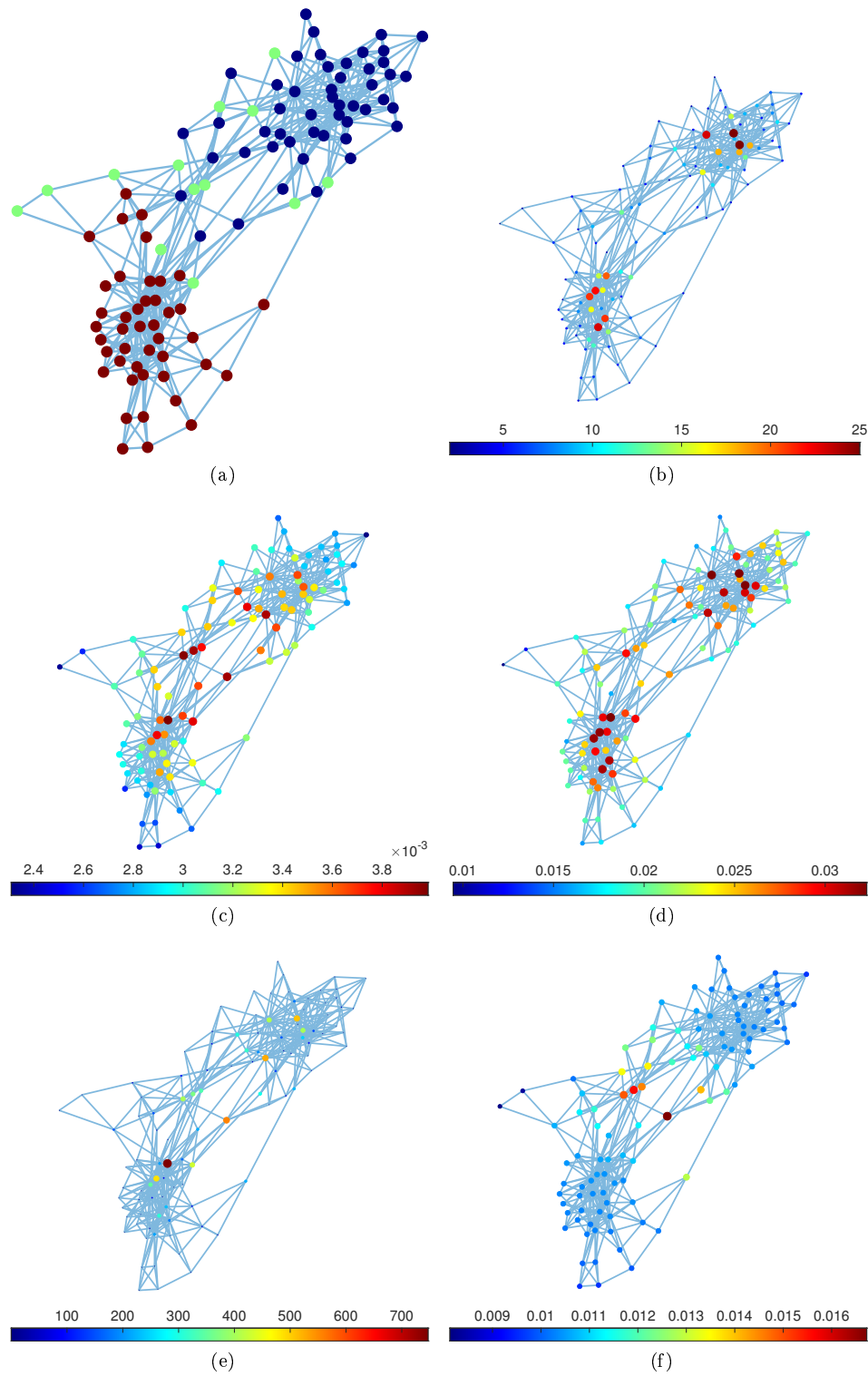


Fig. 6.7: Political books repurchased by the same buyers. (a) A classification of books as liberal (brown), liberal (blue) and neutral (green). (b-f) Coloring of the vertices according to the degree (b), CC (c), RCC (d), betweenness (e) and CCC (f).

Acknowledgements The author thanks financial support from project OLGRA (PID2019-107603GB-I00) funded by Spanish Ministry of Science and Innovation as well as by the Maria de Maeztu project CEX2021-001164-M funded by the MCIN/AEI/10.13039/501100011033. Correspondence with Prof. Chris Godsil is also acknowledged.

Conflict of interest: The author declares that he has no conflict of interest of any kind.

References

1. Wasserman, S., and Faust, K. *Social network analysis: Methods and Applications*. Cambridge University Press, 1994.
2. Watkins, M. E. Automorphisms. In Gross, J. L., Yellen, J., Zhang, P. (eds.) *Handbook of Graph Theory. Second Edition*, CRC Press, Boca Raton, 2014.
3. Bavelas, A. A mathematical model for group structures. *Human organization* 7 (1948) 16-30.
4. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* (1977) 35-41.
5. Freeman, L. C. Centrality in social networks: I. Conceptual clarification. *Social Networks* 1 (1979) 215-239.
6. Estrada, E. *The Structure of Complex Networks: Theory and Applications*. Oxford University Press, 2012.
7. Burago, D., Burago, Y., and Ivanov, S. *A course in metric geometry*. Vol. 33. American Mathematical Society, 2022.
8. Janyasupab, P., Suratanee, A. and Plaimas, K. Network diffusion with centrality measures to identify disease-related genes. *Math. Biosci. Eng* 18.3 (2021) 2909-2929.
9. Saxena, A., and Iyengar, S. Centrality measures in complex networks: A survey. arXiv preprint arXiv:2011.07190 (2020).
10. Rodrigues, F. A. Network centrality: an introduction. In Macau, E. E.N. (Ed.) *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems*. Springer, (2019): 177-196.
11. Chen, D., Cai, D. and Su, H. Self-similarity of complex networks under centrality-based node removal strategy. *Chinese Physics B* (2023).
12. Meghanathan, N. Use of eigenvector centrality to detect graph isomorphism. arXiv preprint arXiv:1511.06620 (2015).
13. Bao, Q., and Zhang, Z. Discriminating power of centrality measures in complex networks. *IEEE Transactions on Cybernetics* 52.11 (2021): 12583-12593.
14. Godsil, C. D. On the full automorphism group of a graph. *Combinatorica* 1 (1981): 243-256.
15. Cameron, P. J., Mary, Q. Automorphisms of graphs. *Topics in algebraic graph theory* 102 (2004): 137-155.
16. Chan, A., and Godsil, C. D. Symmetry and eigenvectors. In: Hahn, G, Sabidussi, G. (Eds). *Graph Symmetry: Algebraic Methods and Applications NATO ASi Series*: (1997): 75-106.
17. Kocay, W. Groups & graphs—software for graphs, digraphs, and their automorphism groups. *MATCH* 58 (2007): 431-443.
18. MacArthur, B. D., Sánchez-García, R. J., Anderson, J. W. Symmetry in complex networks. *Discrete Applied Mathematics* 156.18 (2008): 3525-3531.

19. Sánchez-García, R. J. Exploiting symmetry in network analysis. *Communications Physics* 3.1 (2020): 87.
20. Garlaschelli, D., Ruzzenenti, F., Basosi, R. Complex networks and symmetry I: A review. *Symmetry* 2.3 (2010): 1683-1709.
21. Xiao, Y., MacArthur, B. D., Wang, H., Xiong, M., & Wang, W. (2008). Network quotients: Structural skeletons of complex systems. *Physical Review E*, 78(4), 046102.
22. MacArthur, B.D.; Sanchez-Garcia, R. J. Spectral characteristics of network redundancy. *Phys. Rev. E* 2009, 80, 026117.
23. Koorepazan-Moftakhar, F., and Ashrafi, A. R. Distance under symmetry. *MATCH Commun. Math. Comput. Chem* 74.2 (2015) 259-272.
24. Harary, F., and Palmer, E. D. On similar points of a graph. *Journal of Mathematics and Mechanics* 15.4 (1966): 623-630.
25. Godsil, C. D., and Kocay, W. L. Constructing graphs with pairs of pseudo-similar vertices. *Journal of Combinatorial Theory, Series B* 32.2 (1982): 146-155.
26. Lauri, J. Pseudosimilarity in graphs-a survey. *Ars Combinatoria* 46 (1997): 77-95.
27. Albertson, M. O., Collins, K. L. Symmetry breaking in graphs. *Electr. J. Comb.* 3.1 (1996): R18.
28. Sun, L., Ye, Z., Lu, F., Cui, R., & Wu, C. Similar vertices and isomorphism detection for planar kinematic chains based on ameliorated multi-order adjacent vertex assignment sequence. *Chinese Journal of Mechanical Engineering*, 34 (2021) 1-15.
29. Estrada, E., & Hatano, N. (2008). Communicability in complex networks. *Physical Review E*, 77(3), 036111.
30. Estrada, E., Hatano, N., & Benzi, M. (2012). The physics of communicability in complex networks. *Physics reports*, 514(3), 89-119.
31. Estrada, E., & Rodriguez-Velazquez, J. A. (2005). Subgraph centrality in complex networks. *Physical Review E*, 71(5), 056103.
32. Estrada, E. (2012). The communicability distance in graphs. *Linear Algebra and its Applications*, 436(11), 4317-4328
33. Estrada, E., & Hatano, N. (2016). Communicability angle and the spatial efficiency of networks. *SIAM Review*, 58(4), 692-715.
34. Klein, D. J., & Randić, M. Resistance distance. *Journal of mathematical chemistry*, 12 (1993) 81-95.
35. Bozzo, E., & Franceschet, M. Resistance distance, closeness, and betweenness. *Social Networks*, 35(3) (2013) 460-469.
36. Tarazaga, P., T. L. Hayden, & Wells, J. (1996). Circum-Euclidean distance matrices and faces. *Linear Algebra and its Applications* 232, 77-96.
37. Alfakih, A. Y. (2006) A remark on the faces of the cone of Euclidean distance matrices. *Linear Algebra and its Applications* 414(1), 266-270.
38. Ruhe, A. Perturbation bounds for means of eigenvalues and invariant subspaces. *BIT Numerical Mathematics* 10.3 (1970): 343-354.
39. Estrada, E., Sanchez-Lirola, M. G., & De La Peña, J. A. Hyperspherical embedding of graphs and networks in communicability spaces. *Discrete Applied Mathematics*, 176, (2014) 53-77.
40. Estrada, E. The many facets of the Estrada indices of graphs and networks. *SeMA Journal* 79.1 (2022): 57-125.
41. Estrada, E. Every nonsingular spherical Euclidean distance matrix is a resistance distance matrix. *Linear Algebra and its Applications* 656 (2023): 198-209.

42. Lorrain, F., & White, H. C. Structural equivalence of individuals in social networks. *The Journal of mathematical sociology*, 1(1), (1971) 49-80.
43. Borgatti, S. P., & Everett, M. G. Two algorithms for computing regular equivalence. *Social networks*, 15(4) (1993) 361-376.
44. Batagelj, V., Doreian, P., & Ferligoj, A. An optimizational approach to regular equivalence. *Social networks*, 14(1-2), (1992) 121-135.
45. Blondel, V. D., Gajardo, A., Heymans, M., Senellart, P., & Van Dooren, P. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM review*, 46(4) (2004) 647-666.
46. Kleinberg, J. M. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5) (1999) 604-632.
47. Leicht, E. A., Holme, P., & Newman, M. E. Vertex similarity in networks. *Physical Review E*, 73(2), (2006) 026120.
48. Katz, L. A new status index derived from sociometric analysis. *Psychometrika*, 18(1) (1953) 39-43.
49. *Thesaurus of English Words and Phrases*, edited by S. Mawson T. Y. Crowell Co., New York, 1911.
50. Reitz, J. M. (2004). *Dictionary for library and information science*. Libraries Unlimited.
51. Aguilar, C. O., and Gharesifard, B. Graph controllability classes for the Laplacian leader-follower dynamics. *IEEE Trans. Autom. Control* 60.6 (2014): 1611-1623.
52. Yoon, Myung-Gon, et al. Controllability of multi-agent dynamical systems with a broadcasting control signal. *Asian J. Control* 16.4 (2014): 1066-1072.
53. Rahmani, Amirreza, and Mehran Mesbahi. Pulling the strings on agreement: anchoring, controllability, and graph automorphisms. 2007 American Control Conference. IEEE, 2007.
54. Cvetković, D., Rowlinson, P., Stanić, Z., & Yoon, M. G. Controllable graphs. *Bulletin (Académie serbe des sciences et des arts. Classe des sciences mathématiques et naturelles. Sciences mathématiques)*, (2011) 81-88.
55. Dipert, R. R. The mathematical structure of the world: The world as graph. *The journal of philosophy*, 94(7) (1997) 329-358.
56. Shackel, N. The world as a graph: defending metaphysical graphical structuralism. *Analysis* 71 (2011) 10-21.
57. Sloane, N. J. (2018). *The on-line encyclopedia of integer sequences*. Published electronically at <https://oeis.org>.