Graphs and Combinatorics (2021) 37:493–525
https://doi.org/10.1007/s00373-020-02260-y

ORIGINAL PAPER

Measuring Similarity Between Connected Graphs: The Role of Induced Subgraphs and Complementarity Eigenvalues

Alberto Seeger[1] · David Sossa[2]

Received: 7 July 2020 / Revised: 16 November 2020 / Accepted: 27 November 2020 /
Published online: 2 January 2021

Abstract

This work elaborates on the old problem of measuring the degree of similarity, say $\mathfrak{f}(G, H)$, between a pair of connected graphs G and H, not necessarily of the same order. The choice of a similarity index $\mathfrak{f}$ depends essentially on the graph properties that are considered as important in a given context. As relevant information on a graph, one may consider for instance its degree sequence, its characteristic polynomial, and so on. We explore some new similarity indices based on nonstandard spectral information contained in the graphs under comparison. By nonstandard spectral information in a graph, we mean the set of complementarity eigenvalues of the adjacency matrix. From such a spectral perspective, two distinct graphs G and H are viewed as highly similar if they share a large number of complementarity eigenvalues. This basic idea will be cast in a rigorous mathematical formalism.

Keywords Connected graph · Graph determination · Similarity index · Jaccard's coefficient · Complementarity eigenvalue · Complementarity spectrum · Induced subgraph

Mathematics Subject Classification 05C50 · 15A42

✉ David Sossa
david.sossa@uoh.cl

Alberto Seeger
alberto.seeger@univ-avignon.fr

[1] Department of Mathematics, University of Avignon, 33 rue Louis Pasteur, 84000 Avignon, France

[2] Instituto de Ciencias de la Ingeniería, Universidad de O'Higgins, Av. Libertador Bernardo O'Higgins 611, Rancagua, Chile

1 Introduction

All graphs considered are undirected, unlabeled, loopless and without multiple edges. The attention is further restricted to graphs that are connected. Let $\mathcal{C}$ be the set of connected graphs. The characteristic polynomial $\varphi_G(\lambda) := \det(\lambda I - A_G)$ of a graph G is unambiguously defined because the adjacency matrix A_G is unique up to permutation similarity transformation. As observed in Collatz and Sinogowitz [8], knowing the characteristic polynomial is not enough to determine the graph itself. Two distinct graphs with the same characteristic polynomial are said to be cospectral. Cospectral pairs can be found already among connected graphs of order 6. Cospectral graphs have always the same number of vertices and the same number of edges, but they may not resemble each other; compare for instance Q_1 and Q_2 in Fig. 1. Various substitutes for the characteristic polynomial are proposed in the literature. For instance, Cvetković et al. [10] introduce the generalized characteristic polynomial

$$\phi_G(\lambda, \mu) := \det(\lambda I - A_G + \mu D_G),$$

where D_G stands for the degree matrix of G. Note that ϕ_G is a bivariate polynomial such that $\phi_G(\lambda, 0) = \varphi_G(\lambda)$. Cospectrality relative to ϕ_G is a stronger property than usual cospectrality. Wang et al. [29, Theorem 2.1] show that if two graphs have the same generalized characteristic polynomial, then they resemble each other in at least one important aspect: they have the same degree sequence, cf. Fig. 2.

This result is worthwhile mentioning because the usual characteristic polynomial does not determine the degree sequence. As tool for distinguishing graphs, Liu [17] suggests to use the generalized permanental polynomial

$$\pi_G(\lambda, \mu) := \operatorname{per}(\lambda I - A_G + \mu D_G).$$

Liu [17] asserts that the generalized permanental polynomial is able to distinguish all graphs of order up to 9 and it is also quite efficient for distinguishing graphs of higher order. For instance, there are almost 12 million connected graphs of order 10 and roughly 20% of them have a cospectral mate; by contrast, only a few pairs of connected graphs of order 10 have a common generalized permanental polynomial. Some of these pairs are shown in Fig. 3. Two distinct graphs with the same generalized permanental polynomial are said to be generalized co-permanental. We would like to draw the attention of the reader to the following fact: the generalized permanental polynomial fails to distinguish the connected graphs X_1 and Y_1 in Fig. 3, but these graphs look fairly "similar". The same remark applies to the pairs $\{X_2, Y_2\}$, $\{X_3, Y_3\}$, and $\{X_4, Y_4\}$. In a sense, a possible misidentification made by the

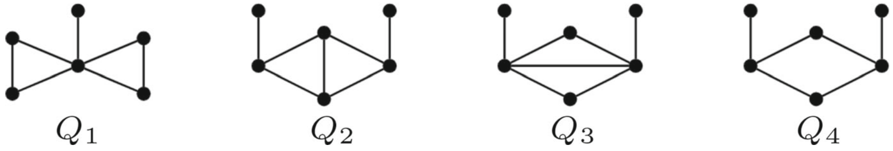

Fig. 1 Q_1 and Q_2 are cospectral, Q_3 and Q_4 have their own characteristic polynomial

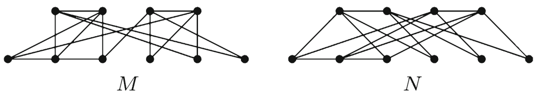

Fig. 2 M and N have the same generalized characteristic polynomial, cf. [29]

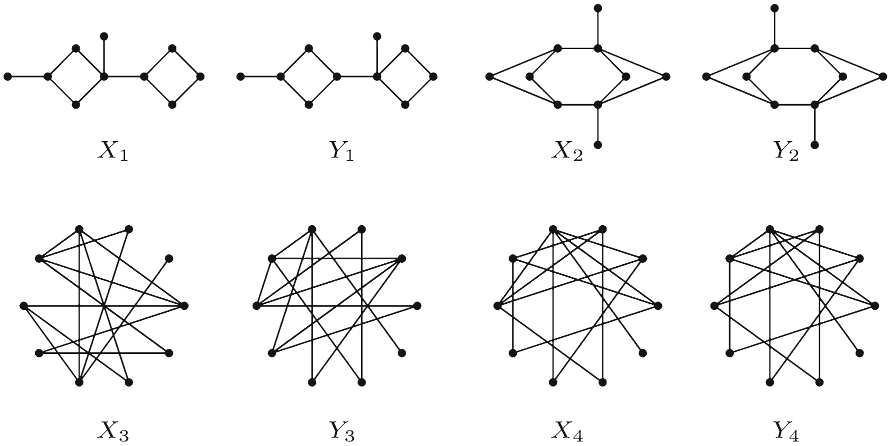

Fig. 3 Pairs of generalized co-permanental graphs, cf. [17]

generalized permanental polynomial is not a too serious mistake. Failing to distinguish X_1 and Y_1 is not so dramatic after all, it would have been worse to misidentify two graphs that look completely different.

1.1 The Concept of Similarity Index

The specific question that we would like to address in this work has to do with the previous discussion on graph determination. It reads as follows: is there a reasonable way of measuring the degree of similarity between two connected graphs, not necessarily of the same order? We allow the graphs to have possibly different orders because such a relaxation gives more flexibility to the theory of similarity indices that we wish to develop. The question under examination concerns the choice of a similarity index on $\mathcal{C}$.

Definition 1 A similarity index on $\mathcal{C}$ is a function $\mathfrak{f} : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ satisfying the following axioms:

$$\begin{aligned} &A_1)\ \mathfrak{f}(G,H) = \mathfrak{f}(H,G)\\ &A_2)\ 0 \leq \mathfrak{f}(G,H) \leq 1\\ &A_3)\ G = H \ \Rightarrow\ \mathfrak{f}(G,H) = 1\\ &A_4)\ \mathfrak{f}(G,H) = 1 \ \Rightarrow\ G = H. \end{aligned}$$

Such an index $\mathfrak{f}$ is of metric type if $\mathfrak{d}(G,H) := 1 - \mathfrak{f}(G,H)$ satisfies the triangle inequality on $\mathcal{C}$.

Recall that we are working with unlabeled graphs. The equality $G = H$ means that G and H are just the same unlabeled graph (of course, if we were to work with labeled graphs, then $G = H$ would mean that G and H are equal up to isomorphism). Informally speaking, that $\mathfrak{f}(G_1, H_1)$ is bigger than $\mathfrak{f}(G_2, H_2)$ means that G_1 is more similar to H_1 than G_2 is to H_2. In short, a higher value of $\mathfrak{f}$ indicates a higher degree of similarity. A metric $\mathfrak{d}$ on $\mathcal{C}$ plays an opposite role of a similarity index: a higher value of $\mathfrak{d}$ indicates a higher degree of dissimilarity. Note however that a similarity index may not be of metric type. Axiom A_2 is not essential but, for the sake of comparison, it is useful to scale numerical values to the standard unit interval [0, 1]. Axioms A_1 and A_3 are quite natural and do not need a further explanation. Axiom A_4 conveys somehow the idea that G and H are highly similar if *f*(*G*, *H*) is near 1. If Axiom A_4 is dropped, then the word index is changed by pseudo-index, in the same way as the word metric is changed by pseudo-metric if the condition $\mathfrak{d}(G, H) = 0$ does not imply $G = H$. When we say that $\mathfrak{f}$ is a similarity criterion, what we mean is that $\mathfrak{f}$ is at least a similarity pseudo-index.

Example 1 A classical way of measuring the distance between two graphs of same order is by means of the expression

$$\mathfrak{d}_*(G, H) := \min_{P \in \mathrm{Perm}(n)} \frac{1}{n} \|P^\top A_G P - A_H\|, \tag{1}$$

where $\|\cdot\|$ is the Frobenius norm, n is the order of both graphs, and $\mathrm{Perm}(n)$ is the set of permutation matrices of order n. The minimum value (1) does not depend on the way of labeling vertices for constructing the adjacency matrices A_G and A_H. The operation $A \mapsto P^\top AP$ is of course a permutation similarity transformation. The optimization problem (1) is known as the graph matching problem, cf. [4, 26]. The scaling factor $1/n$ in front of the Frobenius norm ensures that $\mathfrak{d}_*(G, H)$ is smaller than 1. We write $\mathfrak{d}_*(G, H) = 1$ if G and H are of different order. With such a convention, $\mathfrak{d}_*$ is a metric on the entire set $\mathcal{C}$ and, consequently,

$$\mathfrak{f}_*(G, H) := 1 - \mathfrak{d}_*(G, H) \tag{2}$$

is a similarity index of metric type. Table 1 displays the values of $\mathfrak{f}_*$ on some tests examples. As we can see, the similarity index $\mathfrak{f}_*$ is not sharp enough to discriminate among the different pairs $\{X_k, Y_k\}$. Is this a drawback of $\mathfrak{f}_*$ or is it reasonable to

Table 1 Criterion $\mathfrak{f}_*$ on test examples of order 6 (left) and order 10 (right)

G,H	$\mathfrak{d}_*(G,H)$	$\mathfrak{f}_*(G,H)$	G,H	$\mathfrak{d}_*(G,H)$	$\mathfrak{f}_*(G,H)$
Q_1, Q_2	$\sqrt{8}/6$	0.528596	M, N	$\sqrt{4}/10$	0.800000
Q_1, Q_3	$\sqrt{8}/6$	0.528596	X_1, Y_1	$\sqrt{4}/10$	0.800000
Q_1, Q_4	$\sqrt{10}/6$	0.472954	X_2, Y_2	$\sqrt{4}/10$	0.800000
Q_2, Q_3	$\sqrt{4}/6$	0.666667	X_3, Y_3	$\sqrt{4}/10$	0.800000
Q_2, Q_4	$\sqrt{2}/6$	0.764298	X_4, Y_4	$\sqrt{4}/10$	0.800000
Q_3, Q_4	$\sqrt{2}/6$	0.764298			

consider the pairs $\{X_k, Y_k\}$ as equally similar ? As we shall see later, these pairs are distinguishable by many other similarity criteria.

1.2 Average Similarity and Asymptotic Well-Scaledness

Suppose for a moment that we restrict the attention to a given class $\mathcal{G}$ of connected graphs. To fix the ideas, choose for instance a finite class like

$$\begin{aligned} \mathbb{C}_n &:= \text{connected graphs of order } n \\ \mathbb{T}_n &:= \text{trees of order } n. \end{aligned}$$

Suppose also that we evaluate a certain similarity index $\mathfrak{f}$ on a particular pair $\{G_0, H_0\}$ of graphs taken from the class $\mathcal{G}$. If the numerical evaluation yields for instance $\mathfrak{f}(G_0, H_0) = 0.7$, then should G_0 and H_0 be considered as highly similar graphs or not? For answering this question, it is useful to know the range of values taken by the function $\mathfrak{f}$ on pairs of distinct graphs belonging to $\mathcal{G}$. This leads to consider the optimization problems

$$\begin{aligned} a(\mathfrak{f}, \mathcal{G}) &:= \inf\{\mathfrak{f}(G,H) : \{G,H\} \in \Upsilon(\mathcal{G})\} \\ b(\mathfrak{f}, \mathcal{G}) &:= \sup\{\mathfrak{f}(G,H) : \{G,H\} \in \Upsilon(\mathcal{G})\}, \end{aligned}$$

where $\Upsilon(\mathcal{G})$ is the set of unordered pairs $\{G, H\}$ with $G, H \in \mathcal{G}$ and $G \neq H$. If the class $\mathcal{G}$ under consideration is finite, then we can alternatively use the average value

$$\mu(\mathfrak{f}, \mathcal{G}) := \frac{1}{\operatorname{card}[\Upsilon(\mathcal{G})]} \sum_{\{G,H\} \in \Upsilon(\mathcal{G})} \mathfrak{f}(G,H)$$

as reference level of similarity. By way of example, if $\mu(\mathfrak{f}, \mathcal{G}) = 0.5$, then we can say that G_0 and H_0 have a similarity degree above average in $\mathcal{G}$. By an obvious reason, we would not like the term $\mathfrak{f}(G,H)$ to remain away from 1 as $\{G,H\}$ ranges over all pairs of distinct connected graphs. Analogously, we would not like $\mathfrak{f}(G,H)$ to remain away from 0. It is reasonable, after all, to ask $\mathfrak{f}(G,H)$ to cover as much as possible of the interval [0, 1]. This idea can be formalized as follows.

Definition 2 Let $\mathfrak{f}$ be a similarity index or pseudo-index on $\mathcal{C}$. We say that $\mathfrak{f}$ is asymptotically well-scaled (AWS) if it is both lower-AWS and upper-AWS, i.e.,

$$\lim_{n\to\infty} a(\mathfrak{f}, \mathbb{C}_n) = 0 \quad \text{and} \quad \lim_{n\to\infty} b(\mathfrak{f}, \mathbb{C}_n) = 1, \tag{3}$$

respectively.

The concept of asymptotic well-scaledness on trees is analogous: it suffices to change $\mathbb{C}_n$ by $\mathbb{T}_n$ in (3). Asymptotic well-scaledness on trees implies asymptotic well-scaledness, but not conversely. For instance, the similarity index $\mathfrak{f}_*$ introduced in (2) is AWS but not asymptotically well-scaled on trees. This observation requires perhaps an explanation. A quick computation shows that

$$\begin{aligned} a(\mathfrak{f}_*, \mathbb{C}_n) &\leq \mathfrak{f}_*(P_n, K_n) = 1 - (1/n)\sqrt{(n-1)(n-2)} \\ b(\mathfrak{f}_*, \mathbb{C}_n) &= \mathfrak{f}_*(P_n, C_n) = 1 - (1/n)\sqrt{2}, \end{aligned}$$

where P_n, C_n, and K_n are the path, the cycle, and the complete graph of order n, respectively. We see that $a(\mathfrak{f}_*, \mathbb{C}_n)$ goes to 0 and $b(\mathfrak{f}_*, \mathbb{C}_n)$ goes 1, as needed. Now, if we focus the attention on trees, then the situation is somewhat different. This time we get

$$\begin{aligned} a(\mathfrak{f}_*, \mathbb{T}_n) &= \mathfrak{f}_*(P_n, S_n) = 1 - (1/n)\, 2\sqrt{n-3} \\ b(\mathfrak{f}_*, \mathbb{T}_n) &= \mathfrak{f}_*(P_n, Y_n) = 1 - (1/n)\, 2, \end{aligned}$$

where S_n and Y_n are the star and the snake graph of order n, respectively. By definition, Y_n is obtained by attaching two pendant vertices to a given endvertex of P_{n-2}. Note that both $a(\mathfrak{f}_*, \mathbb{T}_n)$ and $b(\mathfrak{f}_*, \mathbb{T}_n)$ converge to 1. In particular, $\mathfrak{f}_*$ is not lower-AWS on trees. In a sense, $\mathfrak{f}_*$ is ill-scaled for dealing with trees of large order, because we always get values that are near 1. Definition 2 completes the presentation of the mathematical background of this work. Before entering into details, we clarify which are our general goals.

1.3 Scope and Organization of this Work

Finding an optimal permutation matrix P in the graph matching problem (1) is notoriously hard, cf. [4]. The function $\mathfrak{d}_*$ is a graph metric whose numerical evaluation is expensive, even for graphs with only a dozens of vertices. In fact, this remark applies essentially to all graph metrics of interest (graph edit distances [5, 11], Chartrand–Kubicki–Schultz distance [7], etc). We are willing to give up the metric approach and consider any similarity index that is tractable numerically, not necessarily of metric type. We even accept to give up Axiom A_4 and work instead with a similarity pseudo-index, provided such a pseudo-index is based on computable parameters that reflect well the structure of the graph. The choice of parameters (or graph invariants) is of course a crucial point. The organization of the paper is as follows. Undefined terms will be explained in due course.

- In Sect. 2, we introduce and discuss a similarity index, denoted $\mathfrak{f}_{\text{gen}}$, that is based on the family $\mathcal{S}(G)$ of connected induced subgraphs of a given connected graph G. According to this index, two connected graphs G and H are similar if the associated families $\mathcal{S}(G)$ and $\mathcal{S}(H)$ are similar sets in the sense of Jaccard.
- In Sect. 3, we analyze a similarity pseudo-index, denoted $\mathfrak{f}_{\text{spec}}$, that is close in spirit to the previous one. The difference is that now we compare the complementarity spectra $\Pi(G)$ and $\Pi(H)$ of the connected graphs G and H. More precisely, we use Jaccard's coefficient to measure the degree of similarity between these complementarity spectra.
- In Sect. 4, we study two similarity pseudo-indices, denoted $\mathfrak{f}_w$ and $\mathfrak{f}_{\text{lex}}$, that are based on the spectral code of a connected graph G. The spectral code

$$\Gamma(G) := (\varrho_1(G), \varrho_2(G), \varrho_3(G), \ldots)$$

is a sequence that displays the complementarity eigenvalues of G arranged in decreasing order. In the definition of $\mathfrak{f}_{\mathrm{spec}}$, the complementarity eigenvalues of G are considered as equally important; by contrast, while defining $\mathfrak{f}_w$ and $\mathfrak{f}_{\mathrm{lex}}$, the largest complementarity eigenvalue is viewed as more important than the others. Next in importance comes $\varrho_2(G)$, and so on. The idea behind the concept of spectral code of a connected graph is to arrange the complementarity eigenvalues by order of importance.

2 Similarity Indices Based on Induced Subgraphs

2.1 The Genealogical Similarity Index

Let $\mathcal{S}(G)$ be the set of connected induced subgraphs of a connected graph G. Note that $\mathcal{S}(G)$ consists of the graph G itself, together with all its children, all its grandchildren, and so on. By definition, a child of G is an induced subgraph obtained by removing a noncut vertex of G. A grandchild of G is a child of a child of G. The genealogical descendance continues on each branch until ending with a graph of order 1. For economy of language, we refer to $\mathcal{S}(G)$ as the genealogy of G. According to the similarity criterion introduced in the next proposition, two connected graphs G and H are similar if the associated genealogies $\mathcal{S}(G)$ and $\mathcal{S}(H)$ are similar as sets. In what follows, the symbol $\mathcal{M}$ stands for the collection of all nonempty finite sets and $J : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is the function that assigns to a pair (A, B) the value

$$J(A,B) := \frac{\operatorname{card}(A \cap B)}{\operatorname{card}(A \cup B)}. \tag{4}$$

Proposition 1 *An example of similarity index on $\mathcal{C}$ is the function $\mathfrak{f}_{\mathrm{gen}}$ given by*

$$\mathfrak{f}_{\mathrm{gen}}(G,H) := J(\mathcal{S}(G), \mathcal{S}(H)). \tag{5}$$

Such a similarity index is of metric type and AWS.

Proof It is straightforward to check that, for all $A, B \in \mathcal{M}$, we have

$$0 \le J(A,B) = J(B,A) \le 1 = J(A,A).$$

This takes care of Axioms A_1 to A_3. Let $G, H \in \mathcal{C}$ be such that $\mathfrak{f}_{\mathrm{gen}}(G,H) = 1$. Hence, $\mathcal{S}(G) \cap \mathcal{S}(H)$ has the same cardinality as $\mathcal{S}(G) \cup \mathcal{S}(H)$. It follows that $\mathcal{S}(G) = \mathcal{S}(H)$, from where we deduce that $G = H$. So, Axiom A_4 is also in force. That $1 - J$ satisfies the triangle inequality on $\mathcal{M}$ is shown for instance in Levandowski and Winter [16]. This implies that $\mathfrak{d}_{\mathrm{gen}} := 1 - \mathfrak{f}_{\mathrm{gen}}$ satisfies the triangle inequality on $\mathcal{C}$. Finally, a quick computation shows that

$$a(\mathfrak{f}_{\text{gen}}, \mathbb{C}_n) \leq \mathfrak{f}_{\text{gen}}(P_n, K_n) = (n-1)^{-1} \tag{6}$$

$$b(\mathfrak{f}_{\text{gen}}, \mathbb{C}_n) \geq \mathfrak{f}_{\text{gen}}(P_n, C_n) = 1 - 2(n+1)^{-1}. \tag{7}$$

This proves that $\mathfrak{f}_{\text{gen}}$ is AWS. □

By an obvious reason, we call $\mathfrak{f}_{\text{gen}}$ the genealogical similarity index on $\mathcal{C}$ and $\mathfrak{f}_{\text{gen}}(G, H)$ the genealogical similarity degree between G and H. The use of the letter J in the definition of the ratio (4) is not fortuitous: such intersection-to-union ratio was introduced by the botanist Paul Jaccard as a tool for measuring the degree of similarity between two finite sets of possibly different cardinalities. In fact, (4) is known as Jaccard's coefficient (originally given the French name *coefficient de communauté*, cf. [15]). As pointed out in Levandowski and Winter [16], the ratio (4) has a heuristic interpretation: it measures the probability that an element of at least one of the two sets is an element of both, and thus is a reasonable measure of similarity or "overlapping" between the two. For computational convenience, it is sometimes better to write

$$J(A, B) = \frac{\text{card}(A \cap B)}{\text{card}(A) + \text{card}(B) - \text{card}(A \cap B)}$$

as function of the cardinalities of A, B, and their intersection. Consequently, the term (5) reads

$$\mathfrak{f}_{\text{gen}}(G, H) = \frac{r(G, H)}{r(G) + r(H) - r(G, H)}, \tag{8}$$

where $r(G)$ is the cardinality of $\mathcal{S}(G)$ and $r(G, H)$ is the cardinality of the intersection $\mathcal{S}(G) \cap \mathcal{S}(H)$. Let us test the genealogical similarity index on some small order examples.

Example 2 Consider the four graphs in Fig. 1. A handy computation yields

$$\begin{aligned}
\mathcal{S}(Q_1) &= \{\Phi_1, \Phi_2, \Phi_3, \Phi_4, \Phi_6, \Phi_8, \Phi_{18}, \Phi_{21}, Q_1\}\\
\mathcal{S}(Q_2) &= \{\Phi_1, \Phi_2, \Phi_3, \Phi_4, \Phi_5, \Phi_8, \Phi_9, \Phi_{11}, \Phi_{22}, Q_2\}\\
\mathcal{S}(Q_3) &= \{\Phi_1, \Phi_2, \Phi_3, \Phi_4, \Phi_5, \Phi_6, \Phi_8, \Phi_9, \Phi_{17}, \Phi_{23}, Q_3\}\\
\mathcal{S}(Q_4) &= \{\Phi_1, \Phi_2, \Phi_3, \Phi_5, \Phi_6, \Phi_7, \Phi_{11}, \Phi_{15}, Q_4\},
\end{aligned}$$

where the Φ_k's are the connected graphs drawn in the Appendix, cf. Figs. 10 and 11. As one can see, the graphs $Q_1, \ldots, Q_4$, have some connected induced subgraphs in common. For instance, Φ_3 is common to all of them, Φ_9 appears in Q_2 and Q_3, and so on. By evaluating (8), we get

$$
\begin{aligned}
\mathfrak{f}_{\mathrm{gen}}(Q_1,Q_2) &= 5/(9+10-5) = 0.357143\\
\mathfrak{f}_{\mathrm{gen}}(Q_1,Q_3) &= 6/(9+11-6) = 0.428571\\
\mathfrak{f}_{\mathrm{gen}}(Q_1,Q_4) &= 4/(9+\ 9\ -4) = 0.285714\\
\mathfrak{f}_{\mathrm{gen}}(Q_2,Q_3) &= 7/(10+11-7) = 0.500000\\
\mathfrak{f}_{\mathrm{gen}}(Q_2,Q_4) &= 5/(10+9-5) = 0.357143\\
\mathfrak{f}_{\mathrm{gen}}(Q_3,Q_4) &= 5/(11+9-5) = 0.333333.
\end{aligned}
$$

In terms of the criterion $\mathfrak{f}_{\mathrm{gen}}$, the pair $\{Q_2,Q_3\}$ has a higher degree of similarity than the cospectral pair $\{Q_1,Q_2\}$. This may seem strange at first sight, but we must keep in mind that the genealogical similarity index is not intended to be a measure of cospectrality between graphs. The criterion $\mathfrak{f}_{\mathrm{gen}}$ does not focus on characteristic polynomials but on genealogies. The graphs Q_k's belong to the class $\mathbb{C}_6$ of connected graphs of order 6. This class has 112 members and, therefore,

$$\mathrm{card}[\Upsilon(\mathbb{C}_6)] = (1/2)\,112(112-1) = 6216.$$

By evaluating $\mathfrak{f}_{\mathrm{gen}}$ at each one of the 6126 pairs in $\Upsilon(\mathbb{C}_6)$, we get the average value $\mu(\mathfrak{f}_{\mathrm{gen}},\mathbb{C}_6) = 0.382692$, cf. Table 2. Hence, the cospectral pair $\{Q_1,Q_2\}$ has a genealogical similarity degree below average on $\mathbb{C}_6$.

Example 3 Consider the pairs $\{X_1,Y_1\},\ldots,\{X_4,Y_4\}$ shown in Fig. 3. Recall that each $\{X_k,Y_k\}$ is a pair of generalized co-permanental connected graphs of order 10. The pair $\{M,N\}$ shown in Fig. 2 is formed with connected graphs of order 10 with a common generalized characteristic polynomial. Constructing genealogies by hand is quite painful, so we do it with the computer:

$$
\begin{aligned}
\mathfrak{f}_{\mathrm{gen}}(X_1,Y_1) &= 38/(46+44-38) = 0.730769\\
\mathfrak{f}_{\mathrm{gen}}(X_2,Y_2) &= 36/(48+42-36) = 0.666667\\
\mathfrak{f}_{\mathrm{gen}}(X_3,Y_3) &= 75/(108+115-75) = 0.506757\\
\mathfrak{f}_{\mathrm{gen}}(X_4,Y_4) &= 59/(99+97-59) = 0.430657\\
\mathfrak{f}_{\mathrm{gen}}(M,N) &= 53/(87+110-53) = 0.368056.
\end{aligned}
$$

The class $\mathbb{C}_{10}$ of connected graphs of order 10 has 11716571 members. Computing

Table 2 Mean and extreme values of $\mathfrak{f}_{\mathrm{gen}}$ on pairs of connected graphs of order n

n	$a(\mathfrak{f}_{\mathrm{gen}},\mathbb{C}_n)$	$b(\mathfrak{f}_{\mathrm{gen}},\mathbb{C}_n)$	$\mu(\mathfrak{f}_{\mathrm{gen}},\mathbb{C}_n)$
4	0.333333	0.666667	0.497778
5	0.200000	0.777778	0.446368
6	0.142857	0.800000	0.382692
7	0.086957	0.809524	0.321122
8	0.047619	0.850000	0.264879

$\mu(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_{10})$ requires to evaluate $\mathfrak{f}_{\mathrm{gen}}$ on $\mathrm{card}[\Upsilon(\mathbb{C}_{10})] \approx 6.9 \times 10^{13}$ pairs. Such a numerical operation is of course cost prohibitive. For the reader's convenience, we did compute however $\mu(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$ for n up to 8. At the same time, we did compute the extreme values $a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$ and $b(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$, cf. Table 2.

We open a parenthesis and discuss various aspects concerning the information displayed in Table 2. To start with, we comment on the asymptotic behavior of $a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$ as n goes to infinity. Inequality (6) is coarse, but sufficient to prove that $\mathfrak{f}_{\mathrm{gen}}$ is lower-AWS. Table 2 suggests that $a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$ goes to 0 faster than $1/n$. In fact, the next proposition shows that $a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$ goes to 0 faster than any power of $1/n$. In what follows, we use the notation

$$\theta_n := \max_{G \in \mathbb{TF}_n} r(G),$$

where $\mathbb{TF}_n$ stands for the class of triangle-free connected graphs of order n. That a graph is triangle-free means that is does not admits the triangle graph K_3 as induced subgraph.

Proposition 2 *Let* $n \geq 4$. *Then*

$$a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n) \leq 2(\theta_n + n - 2)^{-1} \tag{9}$$

and, in particular, $\lim_{n\to\infty} n^k a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n) = 0$ *for all* $k \in \mathbb{N}$.

Proof Let $n \geq 4$. Clearly,

$$a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n) \leq \min_{G \in \mathbb{C}_n} \mathfrak{f}_{\mathrm{gen}}(G, K_n) \leq \min_{G \in \mathbb{TF}_n} \mathfrak{f}_{\mathrm{gen}}(G, K_n).$$

The last minimization problem can be worked out a bit further. A triangle-free connected graph of order n has only two connected induced subgraphs in common with K_n, namely, K_1 and K_2. On the other hand, it is clear that $r(K_n) = n$. Hence,

$$\mathfrak{f}_{\mathrm{gen}}(G, K_n) = 2(r(G) + n - 2)^{-1}$$

for all $G \in \mathbb{TF}_n$. By passing to the minimum, we get

$$\min_{G \in \mathbb{TF}_n} \mathfrak{f}_{\mathrm{gen}}(G, K_n) = 2(\theta_n + n - 2)^{-1},$$

completing the proof of (9). The integer θ_n corresponds to the maximal number of connected induced subgraphs that can have a triangle-free connected graph of order n. It is very difficult to derive an explicit formula for θ_n. Anyhow, by using the proof of Theorem 6 in Fernandes et al. [13] and the fact that any starlike tree is a triangle-free connected graph, we see that θ_n goes to infinity faster than any polynomial in n. □

Computing θ_n is much cheaper than computing $a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$. The values of θ_n for n until 10 inclusive are displayed in Table 3. The computation of $a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_8)$ took

Table 3 Behavior of θ_n

n	4	5	6	7	8	9	10
$\mathrm{card}(\mathbb{TF}_n)$	3	6	19	59	267	1380	9832
θ_n	4	7	10	18	33	61	122

around 30 h in a computer OS High Sierra, processor 3.4 GHz Intel Core i5 and memory 8 GB. The codes were implemented with Matlab R2017a. We did not compute $a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_9)$, because this would take a month.

The last two columns of Table 4 show that (9) holds as an equality for all n until 7 inclusive, but it is a strict inequality for $n = 8$. The upper bound (9) can be sharpened by adapting the proof of Proposition 2. In fact,

$$\begin{aligned} \min_{G\in\mathbb{C}_n} \mathfrak{f}_{\mathrm{gen}}(G, K_n) &= \min_{G\in\mathbb{C}_n} \frac{\omega(G)}{r(G) + n - \omega(G)} \\ &= \min_{q\in\{2,\ldots,n\}} \frac{q}{\theta_{n,q} + n - q}, \end{aligned} \tag{10}$$

where $\omega(G)$ stands for the clique number of G and

$$\theta_{n,q} := \max_{\substack{G\in\mathbb{C}_n \\ \omega(G)=q}} r(G).$$

In comparison to $2(\theta_n + n - 2)^{-1}$, the minimum (10) is a sharper upper bound for $a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$. However, computing $\theta_{n,q}$ for all $q \in \{2, \ldots, n\}$ is much harder than computing just $\theta_n = \theta_{n,2}$. In praise of (10), we mention that

$$\min_{q\in\{2,\ldots,8\}} \frac{q}{\theta_{8,q} + 8 - q} = \frac{3}{\theta_{8,3} + 8 - 3} = 0.047619 = a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_8),$$

i.e., the bound (10) is optimal at least until $n = 8$.

As far as the asymptotic behavior of $b(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$ is concerned, the lower bound (7) is not sharp but sufficient to prove that $\mathfrak{f}_{\mathrm{gen}}$ is upper-AWS. Obtaining a good estimate for $b(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$ seems a difficult problem. It is not clear to us how to construct a pair $\{G, H\}$ of distinct connected graphs of order n such that $\mathfrak{f}_{\mathrm{gen}}(G, H)$ is as large as possible or nearly as large as possible. Optimal pairs for $n = 7$ and

Table 4 Upper bounds and experimental evaluation of $a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$

n	$(n-1)^{-1}$	$2(\theta_n + n - 2)^{-1}$	$a(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$
4	0.333333	0.333333	0.333333
5	0.250000	0.200000	0.200000
6	0.200000	0.142857	0.142857
7	0.166667	0.086957	0.086957
8	0.142857	0.051282	0.047619
9	0.125000	0.029412	
10	0.111111	0.015385	

$n = 8$ are displayed in Fig. 4. These optimal pairs were obtained by exhaustive numerical testing. For $n = 7$, there are two optimal pairs, but we are displaying only one of them. As we can see from Fig. 4, it is hard to identify the general structure of the graphs forming an optimal pair. Our last comment concerns the asymptotic behavior of the average value $\mu(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$. Table 2 suggests that such average value is decreasing as function of n. It is not clear however if $\mu(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$ goes all the way down to zero.

2.2 The Maximum Common Connected Induced Subgraph Problem

Besides $\mathfrak{f}_{\mathrm{gen}}(G,H)$, there are other similarity indices based on the genealogies of the graphs G and H. For instance, Bunke and Shearer [6] suggest to consider the ratio

$$\mathfrak{f}_{\mathrm{mccis}}(G,H) := \frac{\mathfrak{n}(G,H)}{\max\{|G|,|H|\}}, \tag{11}$$

where $|G|$ stands for the order of G and $\mathfrak{n}(G,H)$ is the largest number of vertices in a connected graph that is an induced subgraph of G and H simultaneously, i.e.,

$$\mathfrak{n}(G,H) := \max_{F \in \mathcal{S}(G) \cap \mathcal{S}(H)} |F|. \tag{12}$$

Note that G and H are not required to be of the same order. A solution F to problem (12) is called a maximum common connected induced subgraph (MCCIS) of the pair $\{G,H\}$. Such an optimal graph F may not be unique, but the term $\mathfrak{n}(G,H)$ is always finite and well defined. Fig. 5 displays a pair of graphs whose MCCIS is unique.

Actually, the maximum common subgraph problem considered in Bunke and Shearer [6] is somewhat different from (12). These authors work with possible disconnected directed graphs and with subgraphs that are not necessarily induced. We are adapting their definition of $\mathfrak{n}(G,H)$ to the context of our work. The ratio (11) is a natural candidate as measure of similarity between connected graphs. Wallis et al. [28] suggest to consider

$$\mathfrak{f}_{\mathrm{wallis}}(G,H) := \frac{\mathfrak{n}(G,H)}{|G| + |H| - \mathfrak{n}(G,H)},$$

an expression that is reminiscent of (8).

Proposition 3 *$\mathfrak{f}_{\mathrm{mccis}}$ and $\mathfrak{f}_{\mathrm{wallis}}$ are similarity indices on $\mathcal{C}$, both of them being AWS.*

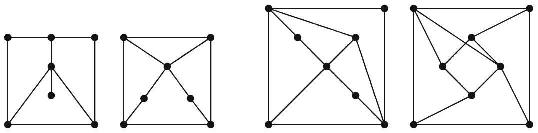

Fig. 4 Pair achieving the maximal value $b(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_n)$ when $n = 7$ (left) and $n = 8$ (right)

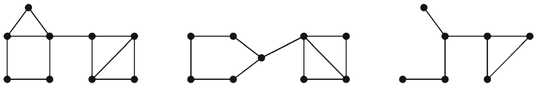

Fig. 5 The third graph is the MCCIS of the first two graphs, cf. [27]

Proof The functions $\tilde{\mathfrak{f}}_{\mathrm{mccis}}$ and $\tilde{\mathfrak{f}}_{\mathrm{wallis}}$ clearly satisfy Axioms A_1 to A_3. For checking that $\tilde{\mathfrak{f}}_{\mathrm{mccis}}$ satisfies Axiom A_4, we consider connected graphs G and H such that $\tilde{\mathfrak{f}}_{\mathrm{mccis}}(G,H) = 1$. Let F be any solution to (12). In such a case,

$$|F| \; = \max\{|G|, |H|\}. \tag{13}$$

Since F is an induced subgraph of both G and H, we have $|F| \leq \min\{|G|, |H|\}$. It follows that $|G| \; = |H| \; = |F|$ and, a posteriori, $G = H = F$. For checking that $\tilde{\mathfrak{f}}_{\mathrm{wallis}}$ satisfies Axiom A_4, we proceed exactly as before, except that (13) must be changed by $|F| \; = (1/2)(|G| + |H|)$. For proving the asymptotic well-scaledness of $\tilde{\mathfrak{f}}_{\mathrm{mccis}}$ and $\tilde{\mathfrak{f}}_{\mathrm{wallis}}$, we simply observe that

$$\begin{aligned}
a(\tilde{\mathfrak{f}}_{\mathrm{mccis}}, \mathbb{C}_n) &= \tilde{\mathfrak{f}}_{\mathrm{mccis}}(P_n, K_n) = 2/n \\
b(\tilde{\mathfrak{f}}_{\mathrm{mccis}}, \mathbb{C}_n) &= \tilde{\mathfrak{f}}_{\mathrm{mccis}}(P_n, C_n) = 1 - (1/n) \\
a(\tilde{\mathfrak{f}}_{\mathrm{wallis}}, \mathbb{C}_n) &= \tilde{\mathfrak{f}}_{\mathrm{wallis}}(P_n, K_n) = (n-1)^{-1} \\
b(\tilde{\mathfrak{f}}_{\mathrm{wallis}}, \mathbb{C}_n) &= \tilde{\mathfrak{f}}_{\mathrm{wallis}}(P_n, C_n) = 1 - 2(n+1)^{-1}.
\end{aligned}$$

This completes the proof of the proposition. □

It is worthwhile mentioning that $\mathfrak{d}_{\mathrm{mccis}} := 1 - \tilde{\mathfrak{f}}_{\mathrm{mccis}}$ and $\mathfrak{d}_{\mathrm{wallis}} := 1 - \tilde{\mathfrak{f}}_{\mathrm{wallis}}$ do not satisfy the triangle inequality on $\mathcal{C}$. To see this, consider for instance the graphs $\{G_1, G_2, G_3\}$ displayed from left to right in Fig. 6. These three graphs are of order 7. Since $\mathfrak{n}(G_1, G_2) = 6$, $\mathfrak{n}(G_2, G_3) = 6$, and $\mathfrak{n}(G_1, G_3) = 4$, we get

$$\begin{aligned}
\underbrace{\mathfrak{d}_{\mathrm{mccis}}(G_1, G_3)}_{3/7} &> \underbrace{\mathfrak{d}_{\mathrm{mccis}}(G_1, G_2)}_{1/7} + \underbrace{\mathfrak{d}_{\mathrm{mccis}}(G_2, G_3)}_{1/7} \\
\underbrace{\mathfrak{d}_{\mathrm{wallis}}(G_1, G_3)}_{3/5} &> \underbrace{\mathfrak{d}_{\mathrm{wallis}}(G_1, G_2)}_{1/4} + \underbrace{\mathfrak{d}_{\mathrm{wallis}}(G_2, G_3)}_{1/4}.
\end{aligned}$$

This example shows that the similarity indices $\tilde{\mathfrak{f}}_{\mathrm{mccis}}$ and $\tilde{\mathfrak{f}}_{\mathrm{wallis}}$ are not of metric type.

Since the average between two numbers cannot exceed their maximum, we readily see that

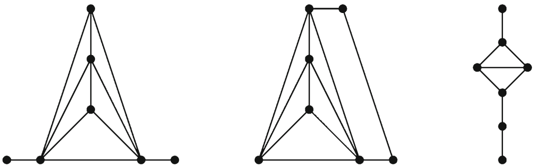

Fig. 6 Counter-example for the triangle inequality in $\mathfrak{d}_{\mathrm{mccis}}$ and $\mathfrak{d}_{\mathrm{wallis}}$

$$\tilde{\mathfrak{f}}_{\rm wallis}(G,H) \geq \left[\frac{2}{\tilde{\mathfrak{f}}_{\rm mccis}(G,H)} - 1\right]^{-1}.$$

If G and H are of the same order, then this inequality becomes an equality and we can write

$$1 + \frac{1}{\tilde{\mathfrak{f}}_{\rm wallis}(G,H)} = \frac{2}{\tilde{\mathfrak{f}}_{\rm mccis}(G,H)}. \tag{14}$$

A major inconvenience of the similarity index $\tilde{\mathfrak{f}}_{\rm mccis}$, or the variant $\tilde{\mathfrak{f}}_{\rm wallis}$, is that the resolution of the maximization problem (12) is a computational nightmare if G and H are not of moderate order. There is a rich literature devoted to various formulations of the maximum common subgraph problem, cf. [9, 20, 27] and references therein. For the sake of illustration, we present below a small order example.

Example 4 Consider the graphs $Q_1,\ldots,Q_4$ in Fig. 1. Since we have already computed each genealogy $\mathcal{S}(Q_k)$, cf. Example 2, it is easy now to see that

$$\begin{aligned}
\mathfrak{n}(Q_1,Q_2) &= |\Phi_8| = 4, \quad \tilde{\mathfrak{f}}_{\rm mccis}(Q_1,Q_2) = 4/6,\\
\mathfrak{n}(Q_1,Q_3) &= |\Phi_8| = 4, \quad \tilde{\mathfrak{f}}_{\rm mccis}(Q_1,Q_3) = 4/6,\\
\mathfrak{n}(Q_1,Q_4) &= |\Phi_6| = 4, \quad \tilde{\mathfrak{f}}_{\rm mccis}(Q_1,Q_4) = 4/6,\\
\mathfrak{n}(Q_2,Q_3) &= |\Phi_9| = 4, \quad \tilde{\mathfrak{f}}_{\rm mccis}(Q_2,Q_3) = 4/6,\\
\mathfrak{n}(Q_2,Q_4) &= |\Phi_{11}| = 5, \quad \tilde{\mathfrak{f}}_{\rm mccis}(Q_2,Q_4) = 5/6,\\
\mathfrak{n}(Q_3,Q_4) &= |\Phi_6| = 4, \quad \tilde{\mathfrak{f}}_{\rm mccis}(Q_3,Q_4) = 4/6.
\end{aligned}$$

The pair $\{Q_1,Q_2\}$ is cospectral, but $\{Q_2,Q_4\}$ is not. Despite this fact, Q_2 is more similar to Q_4 than to Q_1. So, at least with respect to the criterion $\tilde{\mathfrak{f}}_{\rm mccis}$, cospectrality is not a guarantee of maximal degree of similarity.

It is not clear to us whether there is some correlation between $\tilde{\mathfrak{f}}_{\rm mccis}$ and $\tilde{\mathfrak{f}}_{\rm gen}$, but we did observe in numerous examples that $\tilde{\mathfrak{f}}_{\rm mccis}$ is less discriminating than $\tilde{\mathfrak{f}}_{\rm gen}$. Note that $\tilde{\mathfrak{f}}_{\rm mccis}(G,H)$ may take at most $\min\{|G|,|H|\}$ different values, because the order of an optimal solution to (12) is an integer between 1 to $\min\{|G|,|H|\}$ inclusive. If we focus for instance on the class of connected graphs of order 10, then $\tilde{\mathfrak{f}}_{\rm mccis}$ may take at most 10 different values: $0.1, 0.2, \ldots, 0.9, 1$. In fact, the value 0.1 must be ruled out because P_2 is a common connected induced subgraph of any pair of connected graphs of order 10. By way of example, we did solve the maximization problem (12) for each pair $\{X_k,Y_k\}$ in Fig. 3. We used a brute force method which consists in building the genealogies $\mathcal{S}(X_k)$ and $\mathcal{S}(Y_k)$, forming the intersection $\mathcal{S}(X_k)\cap\mathcal{S}(Y_k)$, and selecting a graph of largest order in this intersection. We got

$$\begin{aligned}
\mathfrak{n}(X_1,Y_1) &= 9, \quad \tilde{\mathfrak{f}}_{\rm mccis}(X_1,Y_1) = 0.9\\
\mathfrak{n}(X_2,Y_2) &= 9, \quad \tilde{\mathfrak{f}}_{\rm mccis}(X_2,Y_2) = 0.9\\
\mathfrak{n}(X_3,Y_3) &= 9, \quad \tilde{\mathfrak{f}}_{\rm mccis}(X_3,Y_3) = 0.9\\
\mathfrak{n}(X_4,Y_4) &= 9, \quad \tilde{\mathfrak{f}}_{\rm mccis}(X_4,Y_4) = 0.9
\end{aligned}$$

In a sense, $\tilde{\mathfrak{f}}_{\rm mccis}$ is a rather coarse measure of similarity for connected graphs. The impression of coarseness of $\mathfrak{f}_{\mathrm{mccis}}$ is even more striking in a situation like the following one: consider for instance the class $\mathcal{G}$ of connected graphs of order 6 that admit the path P_5 as induced subgraph. This class has 18 members. If we pick at random two distinct graphs G and H from this class, then we always get $\mathfrak{f}_{\mathrm{mccis}}(G, H) = 5/6$. So, with respect to this class $\mathcal{G}$ at least, the criterion $\mathfrak{f}_{\mathrm{mccis}}$ is totally blind: it does not discriminate among different pairs. Given the relation (14), the same remark applies to the criterion $\mathfrak{f}_{\mathrm{wallis}}$.

3 Similarity Pseudo-Index Based on Complementarity Spectra

As an alternative to the generalized permanental polynomial, we may use the complementarity spectrum as a tool for distinguishing connected graphs. The idea of using complementarity spectra in graph theory is relatively new and was suggested for the first time in Fernandes et al. [13]. Such an idea was further explored in Seeger [22, 23], Seeger and Sossa [24, 25], and Pinheiro et al. [19], just to mention a few published works. The transition from classical eigenvalues to complementarity eigenvalues is a major change in the way of perceiving the concept of spectral information contained in a graph. Recall that a real λ is a complementarity eigenvalue of a graph G if it is a complementarity eigenvalue (or Pareto eigenvalue) of the associated adjacency matrix A_G, i.e., if there exists a nonzero vector $x \in \mathbb{R}^n$ satisfying the complementarity system

$$x \succeq 0,\ A_G x - \lambda x \succeq 0,\ x^\top (A_G x - \lambda x) = 0,$$

where n is the order of G and $x \succeq 0$ means that x is componentwise nonnegative. The complementarity spectrum or set of complementarity eigenvalues of G is denoted by $\Pi(G)$. The set $\Pi(G)$ is nonempty and finite, whether G is connected or not. See Seeger [21] for the theory of complementary spectra of general matrices.

Remark 1 Following a request of one of the referees, we would like to add a few comments on the numerical computation of the complementarity eigenvalues of a symmetric matrix, be it an adjacency matrix or not. If the order of the matrix is moderate, say $n \leq 20$, then Theorem 4.1 in [21] provides an efficient method for detecting all the complementarity eigenvalues: everything boils down to solving a family of classical eigenvalue problems and checking for each one of these problems if the corresponding eigenvectors satisfy a certain system of inequalities. The details can be consulted in [21]. This numerical linear algebra approach is not however the only possibility, one could equally well use a brunch-and-bound optimization algorithm as suggested in [12]. With either method, the computational cost of detecting all the complementarity eigenvalues becomes prohibitive if n goes beyond 30. Nonsmooth Newton type algorithms, as implemented for instance in [1, 2], perform fairly well even if $n = 200$, but, in such a case, it is no longer a realistic aim to capture the entire complementarity spectrum: suffices it to say that a symmetric matrix of order n could have as much as $2^n - 1$ complementarity eigenvalues, cf. [18, Proposition 3].

As shown in Fernandes et al. [13], the complementarity spectrum of a connected graph G admits the representation

$$\Pi(G) = \{\varrho(F) : F \in \mathcal{S}(G)\}, \tag{15}$$

where $\varrho(F)$ stands for the spectral radius of F. In particular, the complementarity spectrum of a connected graph is a nonempty finite subsets of $\mathbb{R}$. The reader should be aware that

$$c(G) := \mathrm{card}[\Pi(G)] \tag{16}$$

may change if G is substituted by another graph of the same order. By way of example, Table 5 displays the complementarity spectra of the graphs of order 6 shown in Fig. 1. For easy of visualization, the complementarity eigenvalues are listed in decreasing order and rounded to 6 decimal places. The second row in Table 5 displays the cardinality of each complementarity spectrum.

Suppose that we wish to measure how similar are the cospectral graphs Q_1 and Q_2. We did consider already the use of the genealogical similarity index $\mathfrak{f}_{\mathrm{gen}}$. Another possibility is to compare the complementarity spectra of Q_1 and Q_2. Note that Q_1 and Q_2 have a certain number of complementarity eigenvalues in common: 2.709275, 2.561553, 2.170086, and so on. There are also some complementarity eigenvalues in one graph but not in the other; for instance, 2.641186 is a complementarity eigenvalue of Q_2 but not of Q_1. A natural way of measuring spectral similarity between Q_1 and Q_2 is to compare the proportion of common complementarity eigenvalues with respect to the number of complementarity eigenvalues of at least one of the two graphs. This leads to the criterion

$$\mathfrak{f}_{\mathrm{spec}}(G,H) := J(\Pi(G),\Pi(H)), \tag{17}$$

where we use again Jaccard's coefficient for quantifying similarity between finite sets.

Table 5 Complementarity spectra of the graphs shown in Fig. 1

Q c(Q)	Q_1 9	Q_2 10	Q_3 11	Q_4 8
ϱ_1	2.709275	2.709275	2.791288	2.236068
ϱ_2	2.561553	2.641186	2.685544	2.135779
ϱ_3	2.342933	2.561553	2.561553	2.000000
ϱ_4	2.170086	2.170086	2.302776	1.732051
ϱ_5	2.000000	2.000000	2.170086	1.618034
ϱ_6	1.732051	1.732051	2.000000	1.414214
ϱ_7	1.414214	1.618034	1.732051	1.000000
ϱ_8	1.000000	1.414214	1.618034	0.000000
ϱ_9	0.000000	1.000000	1.414214	–
ϱ_{10}	–	0.000000	1.000000	–
ϱ_{11}	–	–	0.000000	–

Proposition 4 $\mathfrak{f}_{\text{spec}}$ *is similarity pseudo-index on* $\mathcal{C}$. *Furthermore:*

(a) $\mathfrak{f}_{\text{spec}}$ *is* AWS.
(b) $\mathfrak{d}_{\text{spec}} := 1 - \mathfrak{f}_{\text{spec}}$ *satisfies the triangle inequality on* $\mathcal{C}$.
(c) $\mathfrak{f}_{\text{spec}}$ *is a similarity index on any class of connected graphs on which* Π is injective.

Proof The proof of the first part is as in Proposition 1, except for Axiom A_4. From the very definition of the Jaccard function, it is clear that

$$\mathfrak{f}_{\text{spec}}(G,H) = 1 \quad \Leftrightarrow \quad \Pi(G) = \Pi(H)$$

for all $G, H \in \mathcal{C}$. It is unknown to us whether the function $\Pi : \mathcal{C} \to 2^{\mathbb{R}}$ is injective or not. This issue is quite complex and, despite a considerable effort, we have been unable to clarify this point. Given our current knowledge on the function Π, we can only assert that $\mathfrak{f}_{\text{spec}}$ is a similarity pseudo-index on $\mathcal{C}$. Of course, if $\mathcal{G}$ is any subset of $\mathcal{C}$ on which Π is injective, then $\mathfrak{f}_{\text{spec}}$ is a similarity index on $\mathcal{G}$. The asymptotic well-scaledness of $\mathfrak{f}_{\text{spec}}$ follows from the relations

$$a(\mathfrak{f}_{\text{spec}}, \mathbb{C}_n) \leq \mathfrak{f}_{\text{spec}}(P_n, K_n) = (n-1)^{-1} \tag{18}$$

$$b(\mathfrak{f}_{\text{spec}}, \mathbb{C}_n) \geq \mathfrak{f}_{\text{spec}}(P_n, C_n) = 1 - 2(n+1)^{-1}. \tag{19}$$

For writing the equalities in (18)–(19), we use the known formulas

$$\begin{aligned} \Pi(K_n) &= \{0, 1, 2, \ldots, n-1\} \\ \Pi(P_n) &= \{\beta_1, \ldots, \beta_{n-1}, \beta_n\} \\ \Pi(C_n) &= \{\beta_1, \ldots, \beta_{n-1}, 2\}, \end{aligned}$$

where $\beta_k := 2\cos(\pi/(k+1))$, cf. [22, Example 2.3]. □

We call $\mathfrak{f}_{\text{spec}}$ the spectral similarity pseudo-index on $\mathcal{C}$, because measuring the similarity between two connected graphs G and H is a matter of comparing the complementarity spectra of both graphs. Consistently, we say that $\mathfrak{f}_{\text{spec}}(G,H)$ is the spectral similarity degree between G and H. Written in full extent, the term (17) reads

$$\mathfrak{f}_{\text{spec}}(G,H) = \frac{c(G,H)}{c(G) + c(H) - c(G,H)},$$

where $c(G)$ is defined as in (16) and $c(G, H)$ is the cardinality of $\Pi(G) \cap \Pi(G)$.

Example 5 Let $n \geq 4$ be large. We wish to compare the spectral similarity between the complete graph K_n and the almost complete graph AK_n. Since AK_n is obtained from K_n by removing only one edge, one may naively think that $\mathfrak{f}_{\text{spec}}(K_n, AK_n)$ is near 1. In fact, the deletion of one edge from a graph could change significatively its complementarity spectrum. In the present example,

$$\Pi(K_n) = \{0, 1, 2, \ldots, n-2, n-1\}$$
$$\Pi(AK_n) = \{0, 1, 2, \ldots, n-2\} \cup \{\tau_3, \tau_4, \ldots, \tau_n\},$$

where $\tau_k := \varrho(AK_k)$ is an irrational between $k-2$ and $k-1$. Note that

$$\mathfrak{f}_{\text{spec}}(K_n, AK_n) = \frac{n-1}{n + (2n-3) - (n-1)} = \frac{1}{2}$$

does not go 1 as n goes to infinity.

Example 6 Consider again the four graphs mentioned in Example 2. Table 5 displays the complementarity spectrum of each graph and shows that

$$\mathfrak{f}_{\text{spec}}(Q_1, Q_2) = 8/(9 + 10 - 8) = 0.727273$$
$$\mathfrak{f}_{\text{spec}}(Q_1, Q_3) = 7/(9 + 11 - 7) = 0.538462$$
$$\mathfrak{f}_{\text{spec}}(Q_1, Q_4) = 5/(9 + 8 - 5) = 0.416667$$
$$\mathfrak{f}_{\text{spec}}(Q_2, Q_3) = 8/(10 + 11 - 8) = 0.615385$$
$$\mathfrak{f}_{\text{spec}}(Q_2, Q_4) = 6/(10 + 8 - 6) = 0.500000$$
$$\mathfrak{f}_{\text{spec}}(Q_3, Q_4) = 6/(11 + 8 - 6) = 0.461538.$$

In terms of the criterion $\mathfrak{f}_{\text{spec}}$, the most similar pair is the cospectral pair $\{Q_1, Q_2\}$. Note that the conclusion is not the same as with the criterion $\mathfrak{f}_{\text{gen}}$. On the other hand, by evaluating $\mathfrak{f}_{\text{spec}}$ on each one of the 6126 elements of $\Upsilon(\mathbb{C}_6)$, we get $\mu(\mathfrak{f}_{\text{spec}}, \mathbb{C}_6) = 0.439708$, cf. Table 6. Hence, the cospectral pair $\{Q_1, Q_2\}$ has a spectral similarity above average on $\mathbb{C}_6$.

Example 7 Consider the pairs $\{X_1, Y_1\}, \ldots, \{X_4, Y_4\}$ shown in Fig. 3 and the pair $\{M, N\}$ shown in Fig. 2. A matter of computation shows that

Table 6 Mean and extreme values of $\mathfrak{f}_{\text{spec}}$ on pairs of connected graphs of order n

n	$a(\mathfrak{f}_{\text{spec}}, \mathbb{C}_n)$	$b(\mathfrak{f}_{\text{spec}}, \mathbb{C}_n)$	$\mu(\mathfrak{f}_{\text{spec}}, \mathbb{C}_n)$
4	0.333333	0.800000	0.555556
5	0.222222	0.857143	0.523495
6	0.166667	0.909091	0.439708
7	0.096774	0.923077	0.351142
8	0.052632	0.916667	0.277023

$$\begin{aligned}
\mathfrak{f}_{\text{spec}}(X_1,Y_1) &= 31/(37+36-31) = 0.738095\\
\mathfrak{f}_{\text{spec}}(X_2,Y_2) &= 27/(33+31-27) = 0.729730\\
\mathfrak{f}_{\text{spec}}(X_3,Y_3) &= 70/(96+102-70) = 0.546875\\
\mathfrak{f}_{\text{spec}}(X_4,Y_4) &= 61/(89+89-61) = 0.521368\\
\mathfrak{f}_{\text{spec}}(M,N) &= 49/(70+90-49) = 0.441441.
\end{aligned}$$

Note that X_1 and Y_1 have 31 complementarity eigenvalues in common. This is a big proportion compared to the $37+36-31=42$ complementarity eigenvalues in either one of the graphs. On the other hand, X_3 and Y_3 have 70 complementarity eigenvalues in common. This number is higher than 31, but less important in comparison to $96+102-70=128$.

Let us have a closer look at the asymptotic behavior of $a(\mathfrak{f}_{\text{spec}},\mathbb{C}_n)$ as n goes to infinity. Inequality (18) is coarse, but sufficient to prove that $\mathfrak{f}_{\text{spec}}$ is lower-AWS. It turns out that $a(\mathfrak{f}_{\text{spec}},\mathbb{C}_n)$ goes to 0 faster than any power of $1/n$. This point is explained in the next proposition. Some preliminary words on notation are in order. In general, a complementarity eigenvalue is either an integer or an irrational. This is clear from the representation formula (15) and the fact that the spectral radius of any graph is either an integer or an irrational. In what follows, $\kappa(G)$ denotes the number of integers in the complementarity spectrum of G. Consequently, $c(G)-\kappa(G)$ is the number of irrationals in the complementarity spectrum of G. We also use the notation

$$\mathfrak{c}_n := \max_{G\in\mathbb{C}_n} c(G).$$

The asymptotic behavior of $\mathfrak{c}_n$ has been the object of a long discussion in Seeger and Sossa [25], see also Fernandes et al. [13].

Proposition 5 *Let $n\geq 4$. Then*

$$(\mathfrak{c}_n-1)^{-1} \leq a(\mathfrak{f}_{\text{spec}},\mathbb{C}_n) \leq \xi_n \leq (n-1)(1+\mathfrak{c}_n)^{-1}, \tag{20}$$

where

$$\xi_n := \min_{G\in\mathbb{C}_n} \frac{\kappa(G)}{c(G)-\kappa(G)+n}. \tag{21}$$

In particular, $\lim_{n\to\infty} n^k a(\mathfrak{f}_{\text{gen}},\mathbb{C}_n)=0$ *for all* $k\in\mathbb{N}$.

Proof It is clear that $c(G,H)\geq 2$ for all $G,H\in\mathbb{C}_n$. The first inequality in (20) is because

$$\frac{2}{2\mathfrak{c}_n-2} \leq \frac{2}{c(G)+c(H)-2} \leq \frac{c(G,H)}{c(G)+c(H)-c(G,H)}.$$

The second inequality in (20) is shown as follows. Let $G\in\mathbb{C}_n$. The integers in $\Pi(G)$ are necessarily in $\Pi(K_n)$. Hence, $c(G,K_n)=\kappa(G)$. Computing $\mathfrak{f}_{\text{spec}}(G,K_n)$ is essentially a matter of counting the number of integers and the number of irrationals in the complementarity spectrum of G. Indeed,

$$\mathfrak{f}_{\rm spec}(G, K_n) = \frac{\kappa(G)}{c(G) - \kappa(G) + n}.$$

By passing to the minimum with respect to $G \in \mathbb{C}_n$, we get the desired inequality. The last inequality in (20) is a consequence of the next three observations: firstly, the graph K_n does not attain the minimum in (21); secondly, if $G \in \mathbb{C}_n$ is different from K_n, then $\kappa(G)$ is at most $n - 1$; and, thirdly, the function

$$k \in \{0, 1, \ldots, n-1\} \mapsto (c(G) - k + n)^{-1} k$$

is increasing. For proving the last part of the proposition, we simply recall that $\mathfrak{c}_n$ goes to infinity faster than any polynomial in n, cf. [13]. □

Table 7 shows that, for n until 8 at least, $a(\mathfrak{f}_{\rm spec}, \mathbb{C}_n)$ is equal to ξ_n. We did not compute $a(\mathfrak{f}_{\rm spec}, \mathbb{C}_9)$ because such a computation is too expensive. Seeger and Sossa [25, Section 3] conjecture that $(1/n) \ln \mathfrak{c}_n$ remains away from 0 as n goes to infinity. If such a conjecture is true, then $a(\mathfrak{f}_{\rm spec}, \mathbb{C}_n)$ goes down to 0 at exponential rate. Anyhow, what is clear from (20) is that

$$\lim_{n \to \infty} \frac{\ln(1/a(\mathfrak{f}_{\rm spec}, \mathbb{C}_n))}{\ln \mathfrak{c}_n} = 1,$$

i.e., the logarithms of $1/a(\mathfrak{f}_{\rm spec}, \mathbb{C}_n)$ and $\mathfrak{c}_n$ are asymptotically equivalent.

4 Similarity Pseudo-Indices Based on Spectral Codes

4.1 Inverse-Square and Exponential Decay Versions

The idea behind the concept of spectral code is to arrange the complementarity eigenvalues of a connected graph, say G, by order of importance. To be more precise, the members of the set $\Pi(G)$ are displayed as a sequence

Table 7 Bounds and experimental evaluation of $a(\mathfrak{f}_{\rm spec}, \mathbb{C}_n)$

n	$(\mathfrak{c}_n - 1)^{-1}$	$a(\mathfrak{f}_{\rm spec}, \mathbb{C}_n)$	ξ_n	$(n-1)(1+\mathfrak{c}_n)^{-1}$	$(n-1)^{-1}$
4	0.250000	0.333333	0.333333	0.500000	0.333333
5	0.142857	0.222222	0.222222	0.444444	0.250000
6	0.071429	0.166667	0.166667	0.312500	0.200000
7	0.034483	0.096774	0.096774	0.193548	0.166667
8	0.017857	0.052632	0.052632	0.120690	0.142857
9	0.008333		0.026786	0.065574	0.125000

$$\Gamma(G) := (\varrho_1(G), \varrho_2(G), \varrho_3(G), \ldots) \tag{22}$$

that starts with the largest complementarity eigenvalue of G, then it continues with the second largest, and so on. In general, $\varrho_k(G)$ corresponds to the kth largest complementarity eigenvalue of G, with the convention $\varrho_k(G) = 0$ for all $k \geq c(G)$. The eventually zero sequence (22) is called the spectral code of G. It is reasonable to view two connected graphs G and H as similar if the corresponding spectral codes $\Gamma(G)$ and $\Gamma(H)$ are close to each other. If we were to adopt such an approach, then we could consider for instance a similarity index of the form

$$\mathfrak{f}_{\mathrm{dist}}(G, H) := 1 - \mathrm{dist}[\Gamma(G), \Gamma(H)],$$

where

$$\mathrm{dist}(\gamma, \mu) \; := \; \sum_{k=1}^{\infty} w_k \, \frac{|\gamma_k - \mu_k|}{1 + |\gamma_k - \mu_k|}$$

measures the distance between two eventually zero sequences. The weight or decay factor w_k can be chosen in various ways. Actually, the numerical value of each term $\gamma_k - \mu_k = \varrho_k(G) - \varrho_k(H)$ is somewhat irrelevant to us. What we wish to do in fact is to identify which of these terms are different from zero and count such nonzero terms with a suitable decay factor. This way of proceeding gives an idea of the degree of similarity between the spectral codes $\Gamma(G)$ and $\Gamma(H)$. We propose to examine a similarity criterion of the form

$$\mathfrak{f}_w(G, H) \; := \; \sum_{k=1}^{\infty} w_k \, \chi(\varrho_k(G) - \varrho_k(H)) \tag{23}$$

with

$$\chi(t) := \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{if } t \neq 0. \end{cases}$$

As decay factor in (23) we can use any decreasing sequence $w := \{w_k\}_{k \geq 1}$ of positive numbers such that $\sum_{k=1}^{\infty} w_k = 1$. Two particular instances of (23) are

$$\mathfrak{f}_{\mathrm{ed}}(G, H) := \sum_{k=1}^{\infty} 2^{-k} \, \chi(\varrho_k(G) - \varrho_k(H))$$

$$\mathfrak{f}_{\mathrm{isd}}(G, H) := 6\pi^{-2} \sum_{k=1}^{\infty} k^{-2} \, \chi(\varrho_k(G) - \varrho_k(H)),$$

where the acronyms “ed” and “isd” stand for exponential decay and inverse-square decay, respectively.

Proposition 6 *Let $w := \{w_k\}_{k \geq 1}$ be a decreasing sequence of positive numbers that add up to 1. Then $\mathfrak{f}_w$ is a similarity pseudo-index on $\mathcal{C}$. Furthermore:*

(a) $\mathfrak{f}_w$ *is lower* -AWS.

(b) $\mathfrak{d}_w := 1 - \mathfrak{f}_w$ *satisfies the triangle inequality on* $\mathcal{C}$.
(c) $\mathfrak{f}_w$ *is a similarity index on any class of connected graphs on which* Π *is injective.*

Proof Axioms A_1 and A_3 are obvious. Axiom A_2 is because $\sum_{k=1}^{\infty} w_k = 1$. The discussion concerning (c) is as in Proposition 4. For proving (b), we observe that

$$\mathfrak{d}_w(G,H) = \sum_{k=1}^{\infty} w_k\, z_k(G,H),$$

where each term

$$z_k(G,H) := 1 - \chi(\varrho_k(G) - \varrho_k(H)) = \begin{cases} 0 & \text{if } \ \varrho_k(G) = \varrho_k(H) \\ 1 & \text{if } \ \varrho_k(G) \neq \varrho_k(H) \end{cases}$$

is a $\{0,1\}$ -variable. It is clear that

$$z_k(G,H) \leq z_k(G,F) + z_k(F,H) \tag{24}$$

for all $k \geq 1$. Indeed, if $z_k(G,F) = 0$ and $z_k(F,H) = 0$, then $\varrho_k(G) = \varrho_k(F)$ and $\varrho_k(F) = \varrho_k(H)$, respectively. Hence, $\varrho_k(G) = \varrho_k(H)$ and $z_k(G,H) = 0$. Now, by multiplying each side of (24) by w_k and passing to the sum, we see that $\mathfrak{d}_w$ satisfies the triangle inequality. Finally, we take care of (a). That $\mathfrak{f}_w$ is lower-AWS is simply because

$$a(\mathfrak{f}_w, \mathbb{C}_n) = \mathfrak{f}_w(P_n, Y_n) = 0. \tag{25}$$

The second equality in (25) needs perhaps an explanation. The spectral code of the path P_n is

$$\Gamma(P_n) = (\beta_n, \beta_{n-1}, \ldots, \beta_3, 1, 0, \ldots),$$

where the β_k's are as in the proof of Proposition 4. The spectral code of the snake graph Y_n is also easy to obtain. Indeed, the connected induced subgraphs of Y_n are either paths or smaller order snake graphs:

$$\mathcal{S}(Y_n) = \{P_j : j = 1, \ldots, n-1\} \cup \{Y_i : i = 4, \ldots, n\}.$$

Hence, $\Pi(Y_n) = \Omega_1 \cup \Omega_2$ with

$$\begin{aligned} \Omega_1 &:= \{\varrho(P_j) : j = 1, \ldots, n-1\} = \{\beta_j : j = 1, \ldots, n-1\} \\ \Omega_2 &:= \{\varrho(Y_i) : i = 4, \ldots, n\} = \{\beta_{2i-3} : i = 4, \ldots, n\}. \end{aligned}$$

As observed in the proof of [23, Theorem 2], the sets Ω_1 and Ω_2 intersect each other, but if we let i start from $\lfloor n/2 \rfloor + 2$, then we remove from Ω_2 the elements that are already in Ω_1. Here, $\lfloor \cdot \rfloor$ stands for the lower integer part function. We have identified in this way the

$$c(Y_n) = 2n - \lfloor n/2 \rfloor - 2$$

elements of $\Pi(Y_n)$. For forming the specral code of Y_n, we just need to arrange these elements in decreasing order. By way of example, for $n = 5$ and $n = 6$, we get

$$\begin{cases} \Gamma(P_5) = (\omega_5, \omega_4, \omega_3, 1, 0, \ldots) \\ \Gamma(Y_5) = (\omega_7, \omega_5, \omega_4, \omega_3, 1, 0, \ldots), \end{cases}$$
$$\begin{cases} \Gamma(P_6) = (\omega_6, \omega_5, \omega_4, \omega_3, 1, 0, \ldots) \\ \Gamma(Y_6) = (\omega_9, \omega_7, \omega_5, \omega_4, \omega_3, 1, 0, \ldots), \end{cases}$$

respectively. In general, $\Gamma(P_n)$ and $\Gamma(Y_n)$ share a certain number of terms, but these terms are not placed in the same rank. We see that $\varrho_k(P_n) \neq \varrho_k(Y_n)$ for all $k \geq 1$. Consequently, each $\chi(\varrho_k(P_n) - \varrho_k(Y_n))$ is equal to 0 and

$$\mathfrak{f}_w(P_n, Y_n) = \sum_{k=1}^{\infty} w_k \, \chi(\varrho_k(P_n) - \varrho_k(Y_n)) = 0.$$

Parenthetically, (25) shows not only that $\mathfrak{f}_w$ is lower-AWS, but also lower-AWS on trees. □

Determining whether or not the similarity pseudo-index $\mathfrak{f}_w$ is upper-AWS is a rather difficult question. We shall discuss this issue in tandem with the similarity pseudo-index presented next.

4.2 Lexicographic Version

The first two terms in a spectral code (22) are by far the most important ones. Such terms are easy to compute because they are given by the explicit formulas

$$\varrho_1(G) = \varrho(G), \tag{26}$$

$$\varrho_2(G) = \max_{\substack{F \in \mathcal{S}(G) \\ |F| = |G| - 1}} \varrho(F). \tag{27}$$

We use (26) and (27) if we need to compute only the first two complementarity eigenvalues of a graph, but we rely on Theorem 4.1 in Seeger [21] if we need to compute the entire spectral code. The lexicographic criterion $\mathfrak{f}_{\text{lex}}$ for measuring the degree of similarity between two connected graphs G and H works as follows. We consider the spectral radius as a fundamental parameter of a graph, so we start by computing the spectral radiuses of G and H. If they are different, then we view G and H as highly dissimilar graphs and set $\mathfrak{f}_{\text{lex}}(G, H) = 0$. If G and H have the same spectral radius, then we proceed to compute the second largest complementarity eigenvalue of each graph. If $\varrho_2(G)$ and $\varrho_2(H)$ are different, then we view G and H as partially similar graphs and set for instance $\mathfrak{f}_{\text{lex}}(G, H) = 1/2$. In general, if the truncated spectral codes

$$\Gamma_t(G) := (\varrho_1(G), \varrho_2(G), \ldots, \varrho_t(G))$$
$$\Gamma_t(G) := (\varrho_1(H), \varrho_2(H), \ldots, \varrho_t(H))$$

are equal, then we compare $\varrho_{t+1}(G)$ and $\varrho_{t+1}(H)$ and see whether we continue or stop. What we do in practice is to set

$$\mathfrak{f}_{\text{lex}}(G,H) := 1 - \frac{1}{\mathfrak{s}(G,H)},$$

where the expression

$$\mathfrak{s}(G,H) := \inf\{k : \varrho_k(G) - \varrho_k(H) \neq 0\}$$

has a clear interpretation and, most important, it is computationally tractable. The function $\mathfrak{f}_{\text{lex}}$ enjoys a number of properties.

Proposition 7 $\mathfrak{f}_{\text{lex}}$ *is a similarity pseudo-index on* $\mathcal{C}$. *Furthermore:*

(a) $\mathfrak{f}_{\text{lex}}$ is lower-AWS.
(b) $\mathfrak{d}_{\text{lex}} := 1 - \mathfrak{f}_{\text{lex}}$ satisfies the strong triangle inequality

$$\mathfrak{d}_{\text{lex}}(G,H) \leq \max\{\mathfrak{d}_{\text{lex}}(G,F), \mathfrak{d}_{\text{lex}}(F,H)\} \tag{28}$$

for all $G, F, H \in \mathcal{C}$.
(c) $\mathfrak{f}_{\text{lex}}$ *is a similarity index on any class of connected graphs on which* Π *is injective.*

Proof We just check (28), everything else is now well understood. Let $\text{Seq}(\mathbb{R})$ be the set of sequences $\{\lambda_k\}_{k\geq 1}$ of real numbers. Such a set is known to be a metric space if equipped with the distance function

$$\mathbf{d}(\lambda, \mu) := \frac{1}{\inf\{k \geq 1 : \lambda_k \neq \mu_k\}},$$

where, by convention, the infimum over an empty set is ∞ and $1/\infty$ is equal 0. In fact, it is known that $\mathbf{d}$ satisfies the strong triangle inequality

$$\mathbf{d}(\lambda, \mu) \leq \max\{\mathbf{d}(\lambda, \gamma), \mathbf{d}(\gamma, \mu)\} \tag{29}$$

for all $\lambda, \gamma, \mu \in \text{Seq}(\mathbb{R})$. Said in other words, the set $\text{Seq}(\mathbb{R})$ equipped with the distance function $\mathbf{d}$ is a ultrametric space. Inequality (28) is obtained by applying (29) to the sequences $\lambda = \Gamma(G)$, $\gamma = \Gamma(F)$, and $\mu = \Gamma(H)$. □

For better understanding the similarity pseudo-indices $\mathfrak{f}_{\text{ed}}$, $\mathfrak{f}_{\text{isd}}$, and $\mathfrak{f}_{\text{lex}}$, suppose for a moment that we wish to quantify the degree of similarity between two strings of letters, say

$$G \equiv \text{ johniscalm } \varnothing\ldots,$$
$$H \equiv \text{ johnesclamt } \varnothing\ldots,$$

The symbol $\varnothing$ represents a void space or empty character. Each string is completed with infinitely many void spaces after the first one. The first difference between the strings G and H appears in the 5-th letter ("i" versus "e"). Hence, we set $\mathfrak{f}_{\text{lex}}(G,H) = 1 - (1/5) = 0.800000$, without paying attention to what happens after the 5th letter. The pseudo-index $\mathfrak{f}_{\text{ed}}$ does take into account what happens after the 5-th letter, but the differences observed after the 5-letter are considered less and less important as we move to the right. In this example, we have

$$\begin{aligned}\mathfrak{f}_{\text{ed}}(G,H) &= 2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-6} + 2^{-7} + 2^{-10} + \sum_{k=12}^{\infty} 2^{-k} \\ &= 1 - \left(2^{-5} + 2^{-8} + 2^{-9} + 2^{-11}\right) = 0.962402\end{aligned}$$

The term 2^{-11} in the above line is because "t" is different from the void space. Analogously,

$$\begin{aligned}\mathfrak{f}_{\text{isd}}(G,H) &= 6\pi^{-2}\left(1^{-2} + 2^{-2} + 3^{-2} + 4^{-2} + 6^{-2} + 7^{-2} + 10^{-2} + \sum_{k=12}^{\infty} k^{-2}\right) \\ &= 1 - 6\pi^{-2}\left(5^{-2} + 8^{-2} + 9^{-2} + 11^{-2}\right) = 0.953655\end{aligned}$$

Let us come back to connected graphs. As far as our usual test examples are concerned, the situation is summarized in Table 8.

Table 8 shows that $\mathfrak{f}_{\text{lex}}$ takes the value 0 on non-cospectral pairs. In fact, $\mathfrak{f}_{\text{lex}}(G,H) = 0$ if and only if $\varrho(G) \neq \varrho(H)$. In practice, this equivalence means that the lexicographic criterion $\mathfrak{f}_{\text{lex}}$ is of interest only if we wish to compare the degree of similarity of two connected graphs that have the same spectral radius. If $\varrho(G) = \varrho(H)$, then $\mathfrak{f}_{\text{lex}}(G,H) \geq 1/2$, i.e., $\mathfrak{f}_{\text{lex}}$ jumps from 0 to a value which is at least 1/2. In view of Table 8, it it natural to ask whether $\mathfrak{f}_{\text{lex}}(G,H)$ could be higher than 1/2 for some pair $\{G,H\}$ of distinct connected graphs. This question has a positive answer, but it is not clear to us which is the largest possible value of $\mathfrak{f}_{\text{lex}}(G,H)$. By way of example, consider the pairs $\{Z_1, W_1\}$ and $\{Z_2, W_2\}$ displayed from left to right in Fig. 7.

Table 8 $\mathfrak{f}_{\text{lex}}$, $\mathfrak{f}_{\text{ed}}$, $\mathfrak{f}_{\text{isd}}$ on test examples

G, H	$\mathfrak{s}(G,H)$	$\mathfrak{f}_{\text{lex}}(G,H)$	$\mathfrak{f}_{\text{ed}}(G,H)$	$\mathfrak{f}_{\text{isd}}(G,H)$
Q_1, Q_2	2	0.500000	0.611328	0.751060
Q_1, Q_3	1	0.000000	0.000977	0.057854
Q_1, Q_4	1	0.000000	0.003906	0.071439
Q_2, Q_3	1	0.000000	0.125977	0.125402
Q_2, Q_4	1	0.000000	0.001953	0.063933
Q_3, Q_4	1	0.000000	0.000977	0.057854
X_1, Y_1	1	0.000000	0.250000	0.168636
X_2, Y_2	1	0.000000	0.406250	0.262550
X_3, Y_3	4	0.750000	0.923630	0.896669
X_4, Y_4	2	0.500000	0.703125	0.742860
M, N	4	0.750000	0.875031	0.836951

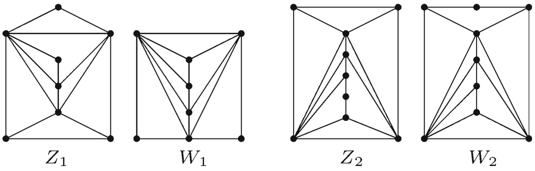

Fig. 7 With respect to the criterion $\mathfrak{f}_{\mathrm{lex}}$, which pair is more similar?

The first pair consists of connected graphs of order 8 and the second pair consists of connected graphs of order 9. Since $\mathfrak{s}(Z_1, W_1) = 9$ and $\mathfrak{s}(Z_2, W_2) = 16$, we have $\mathfrak{f}_{\mathrm{lex}}(Z_1, W_1) = 0.888889$ and $\mathfrak{f}_{\mathrm{lex}}(Z_2, W_2) = 0.937500$. An exhaustive numerical computation shows that the pair $\{Z_1, W_1\} \in \Upsilon(\mathbb{C}_8)$ achieves the maximal value $b(\mathfrak{f}_{\mathrm{lex}}, \mathbb{C}_8)$ and the pair $\{Z_2, W_2\} \in \Upsilon(\mathbb{C}_9)$ achieves the maximal value $b(\mathfrak{f}_{\mathrm{lex}}, \mathbb{C}_9)$. Note that $\mathfrak{f}_{\mathrm{lex}}(Z_2, W_2)$ is not far from 1. If we wish to get closer to 1, say higher than 0.95, then we must work with connected graphs of order 10 at least. As a matter of intensive numerical experimentation, we found a pair $\{Z_3, W_3\}$ of graphs in $\mathbb{C}_{10}$ such that $\mathfrak{s}(Z_3, W_3) = 29$, cf. Fig. 8.

Although $\mathfrak{f}_{\mathrm{lex}}(Z_3, W_3) = 0.965517$ is fairly close to 1, we do not know if the pair $\{Z_3, W_3\}$ achieves the maximal value $b(\mathfrak{f}_{\mathrm{lex}}, \mathbb{C}_{10})$. We did not compute $b(\mathfrak{f}_{\mathrm{lex}}, \mathbb{C}_{10})$ because they are nearly 6.9×10^{13} pairs of connected graphs of order 10. A quick inspection at Table 9 suggests to conjecture that

$$\mathfrak{s}_n := \max_{\{G,H\} \in \Upsilon(\mathbb{C}_n)} \mathfrak{s}(G, H) \tag{30}$$

goes to infinity with n, but we prefer not to advance such a bold statement without a solid justification. The above mentioned conjecture says that $b(\mathfrak{f}_{\mathrm{lex}}, \mathbb{C}_n) = 1 - (1/\mathfrak{s}_n)$ goes to 1 as n goes to infinity, i.e., that $\mathfrak{f}_{\mathrm{lex}}$ is upper-AWS. And what about the similarity pseudo-index $\mathfrak{f}_w$? Is it upper-AWS? Again, this depends on whether or not the term (30) goes to infinity with n. It is not difficult to check that

$$\sum_{k=1}^{\mathfrak{s}(G,H)-1} w_k \leq \mathfrak{f}_w(G, H) \leq 1 - w_{\mathfrak{s}(G,H)} \tag{31}$$

for all $G, H \in \mathcal{C}$. By convention, the sum on the left-hand side of (31) is equal to 0 if $\mathfrak{s}(G, H) = 1$. From here we get

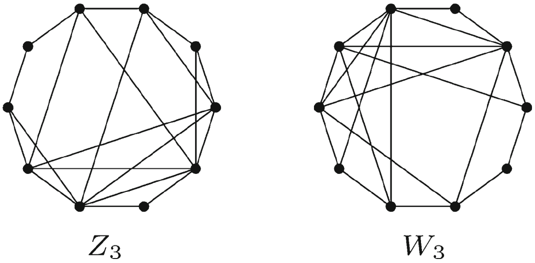

Fig. 8 Graphs sharing the 28 largest complementarity eigenvalues, but not the 29th largest

Table 9 Behavior of $b(\mathfrak{f}_{\text{lex}}, \mathbb{C}_n)$, $b(\mathfrak{f}_{\text{ed}}, \mathbb{C}_n)$, and $b(\mathfrak{f}_{\text{isd}}, \mathbb{C}_n)$

n	4	5	6	7	8	9
$\mathfrak{s}_n$	1	2	2	4	9	16
$b(\mathfrak{f}_{\text{lex}}, \mathbb{C}_n)$	0.000000	0.500000	0.500000	0.750000	0.888889	0.937500
$b(\mathfrak{f}_{\text{ed}}, \mathbb{C}_n)$	0.500000	0.750000	0.750000	0.921883	0.996902	0.999969
$b(\mathfrak{f}_{\text{isd}}, \mathbb{C}_n)$	0.392073	0.848018	0.848018	0.903390	0.956644	

$$\sum_{k=1}^{\mathfrak{s}_n - 1} w_k \leq b(\mathfrak{f}_w, \mathbb{C}_n) \leq 1 - w_{\mathfrak{s}_n}$$

and see that $\mathfrak{f}_w$ is upper-AWS if and only if $\lim_{n\to\infty} \mathfrak{s}_n = \infty$. Unfortunately, it not clear to us which is the behavior of $\mathfrak{s}_n$ as n goes to infinity. It seems that $\mathfrak{s}_n$ increases and goes all the way to infinity, but we do not have a formal proof of this fact.

We mention in passing that (31) yields in particular the estimates

$$6\pi^{-2} \sum_{k=1}^{\mathfrak{s}(G,H)-1} k^{-2} \leq \mathfrak{f}_{\text{isd}}(G,H) \leq 1 - 6\pi^{-2}\, [\mathfrak{s}(G,H)]^{-2} \tag{32}$$

$$1 - (1/2)^{\mathfrak{s}(G,H)-1} \leq \mathfrak{f}_{\text{ed}}(G,H) \leq 1 - (1/2)^{\mathfrak{s}(G,H)}, \tag{33}$$

where the sum on the left-hand side of (32) is taken as 0 if $\mathfrak{s}(G,H) = 1$. As shown in the next proposition, the sandwich (33) can be rewritten as a relation between the similarity pseudo-indices $\mathfrak{f}_{\text{ed}}$ and $\mathfrak{f}_{\text{lex}}$.

Proposition 8 *For any pair $\{G,H\}$ of distinct connected graphs, we have*

$$\frac{\mathfrak{f}_{\text{lex}}(G,H)}{1 - \mathfrak{f}_{\text{lex}}(G,H)} \leq \frac{\ln(1 - \mathfrak{f}_{\text{ed}}(G,H))}{\ln(1/2)} \leq \frac{1}{1 - \mathfrak{f}_{\text{lex}}(G,H)}. \tag{34}$$

For getting (34), it suffices to substitute $\mathfrak{s}(G,H) = (1 - \mathfrak{f}_{\text{lex}}(G,H))^{-1}$ into (33), rearrange terms, and pass to logarithms. Let us return to the discussion on the pairs $\{Z_1, W_1\}$, $\{Z_2, W_2\}$, and $\{Z_3, W_3\}$. If we apply the sandwich (33) to the pair $\{Z_3, W_3\}$, then we get

$$\underbrace{1 - (1/2)^{28}}_{0.9999999963} \leq \mathfrak{f}_{\text{ed}}(Z_3, W_3) \leq \underbrace{1 - (1/2)^{29}}_{0.9999999981}.$$

Note that $\mathfrak{f}_{\text{ed}}$ is nearly constant on pairs $\{G,H\}$ with $\mathfrak{s}(G,H)$ large enough, say $\mathfrak{s}(G,H) \geq 20$. This explains why we are using 10 decimal places in the last entry of Table 10. The criterion $\mathfrak{f}_{\text{ed}}$ is obviously not well adapted to handle such sort of pairs. This problem can be fixed by rescaling $\mathfrak{f}_{\text{ed}}$ in a suitable way; for instance, the exponential decay factor $(1/2)^k$ can be changed by an exponential decay factor like

Table 10 Different ways of measuring similarity in the pairs $\{Z_k, W_k\}$

	Z_1, W_1	Z_2, W_2	Z_3, W_3
$\mathfrak{f}_*$	0.750000	0.685730	0.717157
$\mathfrak{f}_{\mathrm{gen}}$	0.500000	0.467213	0.509091
$\mathfrak{f}_{\mathrm{mccis}}$	0.875000	0.888889	0.900000
$\mathfrak{f}_{\mathrm{wallis}}$	0.777778	0.800000	0.818182
$\mathfrak{f}_{\mathrm{spec}}$	0.600000	0.546296	0.575510
$\mathfrak{f}_{\mathrm{lex}}$	0.888889	0.937500	0.965517
$\mathfrak{f}_{\mathrm{isd}}$	0.956644	0.968627	0.983579
$\mathfrak{f}_{\mathrm{ed}}$	0.996902	0.999969	0.9999999963

$(1/9)(9/10)^k$. Another alternative is to work with the inverse-square decay criterion $\mathfrak{f}_{\mathrm{isd}}$. Note that

$$\underbrace{6\pi^{-2}\sum_{k=1}^{28} k^{-2}}_{0.978671} \;\leq\; \mathfrak{f}_{\mathrm{isd}}(Z_3, W_3) \;\leq\; \underbrace{1 - 6\pi^{-2}\,29^{-2}}_{0.999277}$$

is a more reasonable range of values. Table 10 displays the values of the similarity criteria

$$\underbrace{\mathfrak{f}_*}_{\textit{graph matching}}\;,\;\underbrace{\overbrace{\mathfrak{f}_{\mathrm{gen}}}^{\textit{genealogy}}\;,\;\overbrace{\mathfrak{f}_{\mathrm{mccis}}\;,\;\mathfrak{f}_{\mathrm{wallis}}}^{\textit{common subgraph}}}_{\textit{induced subgraphs}}\;,\;\underbrace{\overbrace{\mathfrak{f}_{\mathrm{spec}}}^{\textit{compl. spectrum}}\;,\;\overbrace{\mathfrak{f}_{\mathrm{lex}},\,\mathfrak{f}_{\mathrm{isd}},\,\mathfrak{f}_{\mathrm{ed}}}^{\textit{spectral code}}}_{\textit{spectral information}}$$

on the pairs $\{Z_k, W_k\}$. For pedagogical reasons, we are partitioning our collection of similarity criteria into different classes and subclasses.

Table 10 shows that $\{\mathfrak{f}_{\mathrm{lex}}, \mathfrak{f}_{\mathrm{isd}}, \mathfrak{f}_{\mathrm{ed}}\}$ take values near 1 on each $\{Z_k, W_k\}$. This observation is consistent with the fact that $\{\mathfrak{f}_{\mathrm{lex}}, \mathfrak{f}_{\mathrm{isd}}, \mathfrak{f}_{\mathrm{ed}}\}$ take values near 1 on any pair $\{G, H\}$ such that $\mathfrak{s}(G, H)$ is large. In fact, our three similarity criteria based on the comparison of spectral codes have been constructed on purpose so that $\mathfrak{f}(G, H)$ is near 1 whenever $\mathfrak{s}(G, H)$ is large. The motivation behind the construction of the other similarity criteria is somewhat different, but we observe that $\mathfrak{f}(Z_k, W_k)$ is still high for other choices of $\mathfrak{f}$. For instance,

$$\mathfrak{f}_{\mathrm{mccis}}(Z_3, W_3) = 0.900000 = \; b(\mathfrak{f}_{\mathrm{mccis}}, \mathbb{C}_{10}),$$

i.e., $\{Z_3, W_3\}$ is a pair of connected graphs of order 10 that achieves the value $b(\mathfrak{f}_{\mathrm{mccis}}, \mathbb{C}_{10})$. Although

$$\mathfrak{f}_{\mathrm{gen}}(Z_3, W_3) = \frac{140}{210 + 205 - 140} = 0.509091$$

is much lower than the maximal value $b(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_{10})$, it is nonetheless well above the average value $\mu(\mathfrak{f}_{\mathrm{gen}}, \mathbb{C}_{10})$. Note that Z_3 and W_3 have 140 connected induced subgraphs in common. On the other hand, there are 275 connected induced subgraphs of either Z_3 or W_3, but not of both graphs. The presence of these 275 non-common induced subgraphs explains why $\mathfrak{f}_{\mathrm{gen}}(Z_3, W_3)$ is not that large.

Remark 2 It was hard to detect a pair $\{Z_3, W_3\}$ with $\mathfrak{s}(Z_3, W_3)$ as large as 29. Most likely there is another pair, say $\{Z_4, W_4\}$, with $\mathfrak{s}(Z_4, W_4)$ equal to 30 or higher, but finding such a pair requires an heavy computational investment. In any case, we have no theoretical strategy that helps in detecting such a pair $\{Z_4, W_4\}$.

5 Conclusions

The criteria $\mathfrak{f}_{\mathrm{gen}}$ and $\mathfrak{f}_{\mathrm{spec}}$ have something in common: both of them can be expressed as

$$\mathfrak{f}(G, H) := J(\sigma(G), \sigma(H)), \tag{35}$$

where $\sigma : \mathcal{C} \to \mathcal{M}$ is function that converts a connected graph into a nonempty finite set. The role of the set $\sigma(G)$ is to gather relevant information on the structure of G (for instance, the induced subgraphs of G, the complementarity eigenvalues of G, etc). According to (35), the task of measuring the degree of similarity between two connected graphs G and H is a matter of evaluating the Jaccard coefficient between $\sigma(G)$ and $\sigma(H)$. As an alternative to (35), we could perfectly well consider a criterion

$$\mathfrak{f}^{\circ}(G, H) := J^{\circ}(\sigma(G), \sigma(H)),$$

in which Jaccard's coefficient is changed by Sørensen-Dice's coefficient

$$J^{\circ}(A, B) := \frac{2\,\mathrm{card}(A \cap B)}{\mathrm{card}(A) + \mathrm{card}(B)}. \tag{36}$$

The ratio (36) is yet another way of measuring the degree of similarity between a pair of finite sets. Such a ratio was introduced independently by the botanists Thorvald Sørensen and Lee Raymond Dice in 1948 and 1945 respectively. It turns out that J° and J are equivalent in the sense that

$$J^{\circ}(A, B) = \psi(J(A, B))$$

for some increasing ψ mapping [0, 1] onto [0, 1], namely, $\psi(t) := 2t/(1 + t)$. So, there is no loss of generality in focusing on Jaccard's coefficient. The important question is not whether we use J or J°, but whether we choose $\sigma(G) = \mathcal{S}(G)$ or $\sigma(G) = \Pi(G)$ as set representing the structure of G.

Of all the similarity criteria mentioned in Table 11, which one is the best? This is a question whose answer depends on the context and, in any case, it remains subject to interpretation. There is no universal agreement on how the expression "similarity between graphs" is to be understood. We propose an axiomatic approach for handling the concept of similarity: in order to measure the degree of similarity between two connected graphs, we suggest to use a similarity index or, more generally, a similarity pseudo-index. Similarity indices and similarity pseudo-

Table 11 Properties of various similarity indices and/or pseudo-indices

	Metric type	Axioms		AWS	
		A_1-A_3	A_4	Lower	Upper
$\mathfrak{f}_*$	✔	✔	✔	✔	✔
$\mathfrak{f}_{\mathrm{gen}}$	✔	✔	✔	✔	✔
$\mathfrak{f}_{\mathrm{mccis}}$	×	✔	✔	✔	✔
$\mathfrak{f}_{\mathrm{wallis}}$	×	✔	✔	✔	✔
$\mathfrak{f}_{\mathrm{spec}}$	✔	✔	?	✔	✔
$\mathfrak{f}_{\mathrm{ed}}$	✔	✔	?	✔	?
$\mathfrak{f}_{\mathrm{isd}}$	✔	✔	?	✔	?
$\mathfrak{f}_{\mathrm{lex}}$	✔	✔	?	✔	?

indices are functions $\mathfrak{f} : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ satisfying certain axioms, cf. Definition 1. There are at least three interesting ways of constructing the bivariate function $\mathfrak{f}$, namely,

- $\mathfrak{f}(G, H)$ is a number that depends on the genealogies $\mathcal{S}(G)$ and $\mathcal{S}(H)$; for instance, the similarity indices $\mathfrak{f}_{\mathrm{gen}}$, $\mathfrak{f}_{\mathrm{mccis}}$, and $\mathfrak{f}_{\mathrm{wallis}}$, belong to this category.
- $\mathfrak{f}(G, H)$ is a number that depends on the complementarity spectra $\Pi(G)$ and $\Pi(H)$; an example of this case is the similarity pseudo-index $\mathfrak{f}_{\mathrm{spec}}$. The rationale behind the definition of $\mathfrak{f}_{\mathrm{spec}}$ is that the complementarity spectrum of a graph represents well the graph under consideration.
- $\mathfrak{f}(G, H)$ is a number that depends on the spectral codes $\Gamma(G)$ and $\Gamma(H)$; for instance, the similarity pseudo-indices $\mathfrak{f}_{\mathrm{lex}}$, $\mathfrak{f}_{\mathrm{ed}}$, and $\mathfrak{f}_{\mathrm{isd}}$, fit into this model.

Should we give the priority to genealogies, to complementarity spectra, or to spectral codes? In fact, each choice has its own advantages and disadvantages. From a practical point of view, it is helpful to keep in mind the following three facts. Firstly, for a connected graph G of arbitrary order but with special structure (broom, complete bipartite graph, starlike tree, etc), it is easier to construct the genealogy $\mathcal{S}(G)$ than to compute the complementarity spectrum $\Pi(G)$. Secondly, for a randomly generated (or unstructured) connected graph G, it is cheaper to compute $\Pi(G)$ than $\mathcal{S}(G)$. For computing $\Pi(G)$ we do not need to bother with the problem of testing connectedness in each induced subgraph, nor with the expensive task of testing whether two induced subgraphs are isomorphic (when represented as labeled graphs). As it is well known, evaluating the cardinality of $\mathcal{S}(G)$ is an old and difficult problem of graph theory, cf. [3, 13, 14]. And, thirdly, computing $\Gamma(G)$ is the same cost as computing $\Pi(G)$. The only difference between both mathematical objects is that $\Gamma(G)$ contains the elements of $\Pi(G)$ is a prescribed order: largest elements comes first. The difference between $\Gamma(G)$ and $\Pi(G)$ is conceptual, not computational.

From a theoretical point of view, it may be of interest to ask $\mathfrak{f}$ to satisfy not just the axioms of a similarity pseudo-index, but also some additional properties: being of metric type and being ASW are of course two favorable properties. Table 11 summarizes our current knowledge on the different similarity criteria considered in this work.

A question mark in Table 11 simply means that we do not know how to fill the corresponding place. The four question marks under the column A_4 are because we do not know whether the complementarity spectrum of a connected graph is enough to determine the graph itself. Such an issue is very difficult to settle and remains unsolved. As a partial answer obtained by exhaustive numerical experimentation, we mention that the implication

$$\Pi(G) = \Pi(H) \quad \Rightarrow \quad G = H \tag{37}$$

is true at least if G and H are connected graphs on fewer than 10 vertices. Implication (37) is also true if G and H belong to some special classes of highly structured connected graphs. This specific theme is the object of a work of ours that is still in progress.

We close this work with a few words concerning the link between cospectrality and graph similarity. It was hinted in Sect. 1 that cospectral graphs may not resemble each other. From our large battery of numerical experiments, we have learned that a cospectral pair $\{G, H\}$ can be highly dissimilar, even if the similarity criterion $\mathfrak{f}$ under consideration is reasonable and intuitive. For instance, as seen in Example 2, the cospectral pair $\{Q_1, Q_2\}$ has a genealogical similarity degree below average. Along the same lines, it is possible to construct cospectral pairs that are highly dissimilar with respect to $\mathfrak{f}_{\text{spec}}$. Specially striking examples are the cospectral pairs $\{Q_5, Q_6\}$ and $\{Q_7, Q_8\}$ displayed from left to right in Fig. 9.

The first cospectral pair has low genealogical similarity degree, namely, $\mathfrak{f}_{\text{gen}}(Q_5, Q_6) = 0.192308$, whereas these second cospectral pair has low spectral similarity degree, namely, $\mathfrak{f}_{\text{spec}}(Q_7, Q_8) = 0.260000$. Table 12 displays the minimal value

$$e(\mathfrak{f}, \mathbb{C}_n) := \min\{\mathfrak{f}(G, H) : G, H \in \mathbb{C}_n \ \text{ with } \{G, H\} \text{ cospectral } \}$$

when the similarity criterion $\mathfrak{f}$ is either $\mathfrak{f}_{\text{gen}}$ or $\mathfrak{f}_{\text{spec}}$. We let n range from 6 to 8 inclusive, going beyond would take too much computational time.

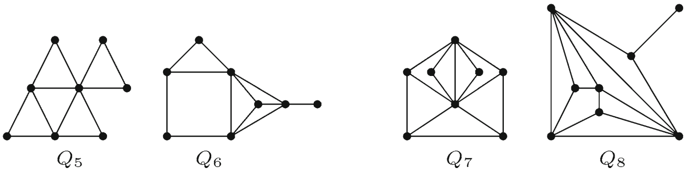

Fig. 9 Cospectral pairs with low genealogical and spectral similarity degrees, respectively

Table 12 $e(\mathfrak{f}_{\text{gen}}, \mathbb{C}_n)$ and $e(\mathfrak{f}_{\text{spec}}, \mathbb{C}_n)$

n	6	7	8
$e(\mathfrak{f}_{\text{gen}}, \mathbb{C}_n)$	0.357143	0.241379	0.192308
$e(\mathfrak{f}_{\text{spec}}, \mathbb{C}_n)$	0.727273	0.375000	0.260000

Appendix

Connected graphs can be enumerated in manifold ways. This work uses the enumeration system $\{\Phi_k\}_{k\geq 1}$ introduced in Seeger [22]. There is no need of explaining here the rationale behind such an enumeration technique. Suffices it to say that one starts with the small graphs $\Phi_1 = K_1$, $\Phi_2 = K_2$, $\Phi_3 = P_3$, and $\Phi_4 = K_3$. Then one continues with the connected graphs on 4 vertices, arranged in increasing order according to the spectral radius, cf. Fig. 10.

Next comes the graphs on 5 vertices, also arranged in increasing order according to the spectral radius. Since C_5 and S_5 have the same spectral radius, the graph with smallest second largest complementarity eigenvalue comes first. Figure 11 shows the connected graphs on 5 vertices mentioned in this work.

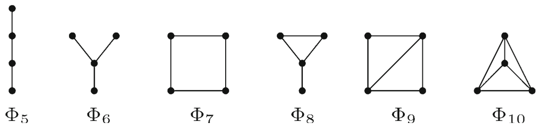

Fig. 10 Connected graphs on 4 vertices, arranged by increasing spectral radius

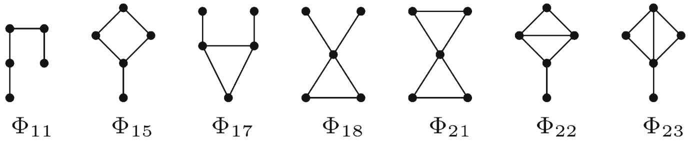

Fig. 11 Connected graphs on 5 vertices mentioned in this work

References

1. Adly, S., Rammal, H.: A new method for solving Pareto eigenvalue complementarity problems. Comput. Optim. Appl. **55**, 703–731 (2013)
2. Adly, S., Seeger, A.: A nonsmooth algorithm for cone-constrained eigenvalue problems. Comput. Optim. Appl. **49**, 299–318 (2011)
3. Alon, N., Bollobás, B.: Graphs with a small number of distinct induced subgraphs. Discrete Math. **75**, 23–30 (1989)

4. Bento, J., Ioannidis, S.: A family of tractable graph metrics. Appl. Netw. Sci. **4**, 107 (2019). https://doi.org/10.1007/s41109-019-0219-z
5. Bougleux S., Gaüzère B., Brun L.: Graph edit distance as a quadratic program. In: ICPR 2016, 23rd International Conference on Pattern Recognition, Cancun, hal:01418937 (2016)
6. Bunke, H., Shearer, K.: A graph distance metric based on the maximal common subgraph. Pattern Recognit. Lett. **19**, 255–259 (1998)
7. Chartrand, G., Kubicki, G., Schultz, M.: Graph similarity and distance in graphs. Aequ. Math. **55**, 129–145 (1998)
8. Collatz, L., Sinogowitz, U.: Spektren endlicher Grafen. Abh. Math. Sem. Univ. Hamburg **21**, 63–77 (1957)
9. Cuissart B., Hébrard J.-J.: A direct algorithm to find a largest common connected induced subgraph of two graphs. In: Brun, L., Vento, M. (eds.) Proc. 5th Int. Workshop on Graph-based Representations in Pattern Recognition, LNCS 3434, pp. 162–171. Springer (2005)
10. Cvetković, D.M., Doob, M., Sachs, H.: Spectra of Graphs. Academic, New York (1982)
11. Gao, X., Xiao, B., Tao, D., Li, X.: A survey of graph edit distance. Pattern Anal. Appl. **13**, 113–129 (2010)
12. Fernandes, L., Júdice, J., Sherali, H., Fukushima, M.: On the computation of all eigenvalues for the eigenvalue complementarity problem. J. Glob. Optim. **59**, 307–326 (2014)
13. Fernandes, R., Júdice, J., Trevisan, V.: Complementarity eigenvalue of graphs. Linear Algebra Appl. **527**, 216–231 (2017)
14. Harary, F., Palmer, E.M.: Graphical Enumeration. Academic Press, New York (1973)
15. Jaccard, P.: Etude comparative de la distribution florale dans une portion des Alpes et des Jura. Bull. Soc. Vaudoise Sciences Naturelles. **37**, 547–579 (1901)
16. Levandowsky, M., Winter, D.: Distance between sets. Nature **234**, 34–35 (1971)
17. Liu, S.: Generalized permanental polynomials of graphs. Symmetry **11**(2), 242 (2019). https://doi.org/10.3390/sym11020242
18. Pinto da Costa, A., Seeger, A.: Cone-constrained eigenvalue problems: theory and algorithms. Comput. Optim. Appl. **45**, 25–57 (2010)
19. Pinheiro, L.K., Souza, B.S., Trevisan, V.: Determining graphs by the complementary spectrum. Discrete Math. Graph Theory **40**, 607–620 (2020)
20. Raymond, J.W., Willet, P.: Maximum common subgraph isomorphism algorithms for the matching of chemical structures. J. Comput. Aided Mol. Des. **16**, 521–533 (2002)
21. Seeger, A.: Eigenvalue analysis of equilibrium processes defined by linear complementarity conditions. Linear Algebra Appl. **292**, 1–14 (1999)
22. Seeger, A.: Complementarity spectral analysis of connected graphs. Linear Algebra Appl. **543**, 205–225 (2018)
23. Seeger, A.: Repetition of spectral radiuses among connected induced subgraphs. Graphs Comb. **36**, 1131–1144 (2020)
24. Seeger, A., Sossa, D.: Extremal problems involving the two largest complementarity eigenvalues of a graph. Graphs Comb. **36**, 1–25 (2020)
25. Seeger, A., Sossa, D.: On cardinality of complementarity spectra of connected graphs. Linear Algebra Appl. (2019). https://doi.org/10.1016/j.laa.2019.11.012
26. Umeyama, S.: An eigen decomposition approach to weighted graph matching problems. IEEE Trans. Pattern Anal. Mach. Intell. **10**, 695–703 (1988)
27. Vismara, P., Valery, B.: Finding maximum common connected subgraphs using clique detection or constraint satisfaction algorithms. In: Le Thi, H.A., Bouvry, P., Pham Dinh, T. (eds.) Modelling, Computation and Optimization in Information Systems and Management Sciences, pp. 358–368. Springer, Berlin (2008)
28. Wallis, W.D., Shoubridge, P., Kraetz, M., Ray, D.: Graph distances using graph union. Pattern Recognit. Lett. **22**, 701–704 (2001)
29. Wang, W., Li, F., Lu, H., Xu, Z.: Graphs determined by their generalized characteristic polynomials. Linear Algebra Appl. **434**, 1378–1387 (2011)