Yugoslav Journal of Operations Research
22 (2012), Number 2, 145-161
DOI: 10.2298/YJOR120925025C

SPECTRAL RECOGNITION OF GRAPHS[1]

Dragoš CVETKOVIĆ

Mathematical Institute SANU,
P.O. Box 367, 11000 Belgrade, Serbia
ecvetkod@etf.rs

Invited review

Abstract: At some time, in the childhood of spectral graph theory, it was conjectured that non-isomorphic graphs have different spectra, i.e. that graphs are characterized by their spectra. Very quickly this conjecture was refuted and numerous examples and families of non-isomorphic graphs with the same spectrum (cospectral graphs) were found. Still some graphs are characterized by their spectra and several mathematical papers are devoted to this topic.

In applications to computer sciences, spectral graph theory is considered as very strong. The benefit of using graph spectra in treating graphs is that eigenvalues and eigenvectors of several graph matrices can be quickly computed. Spectral graph parameters contain a lot of information on the graph structure (both global and local) including some information on graph parameters that, in general, are computed by exponential algorithms. Moreover, in some applications in data mining, graph spectra are used to encode graphs themselves.

The Euclidean distance between the eigenvalue sequences of two graphs on the same number of vertices is called the spectral distance of graphs. Some other spectral distances (also based on various graph matrices) have been considered as well. Two graphs are considered as similar if their spectral distance is small. If two graphs are at zero distance, they are cospectral. In this sense, cospectral graphs are similar. Other spectrally based measures of similarity between networks (not necessarily having the same number of vertices) have been used in Internet topology analysis, and in other areas.

The notion of spectral distance enables the design of various meta-heuristic (e.g., tabu search, variable neighbourhood search) algorithms for constructing graphs with a given spectrum (spectral graph reconstruction).

---

[1] Presented at the Conference on Applications of Graph Spectra in Computer Science, Barcelona, July 16 to 20, 2012

Several spectrally based pattern recognition problems appear in many areas (e.g., image segmentation in computer vision, alignment of protein-protein interaction networks in bio-informatics, recognizing hard instances for combinatorial optimization problems such as the travelling salesman problem).

We give a survey of such and other graph spectral recognition techniques used in computer sciences.

Keywords: Spectral graph theory, spectral recognition, computer science, internet, data mining.

MSC: 05C50, 68P20, 68R10.

1. INTRODUCTION

Spectral graph theory is a mathematical theory in which linear algebra and graph theory meet. For any graph matrix M we can build a spectral graph theory in which graphs are studied by means of eigenvalues of the matrix M. This theory is called M-theory. In order to avoid confusion, to any notion in this theory a prefix M-could be added (e.g., M-eigenvalues). Frequently used graph matrices are the adjacency matrix A, the Laplacian $L = D - A$ and the signless Laplacian $Q = D + A$, where D is a diagonal matrix of vertex degrees. *The* spectral graph theory includes all particular theories together with interaction tools.

1.1. Preliminaries: Spectral graph theory in computer sciences

It was recognized in about the last ten years that graph spectra have several important applications in computer sciences (see, e.g., [13, 14, 21]). Graph spectra appear in the literature in Internet technologies, computer vision, pattern recognition, data mining, multiprocessor systems, statistical databases and in many other areas. There are thousands of such papers.

In surveys [21] and [2] of the applications of graph spectra in Computer science, applications in the following branches of Computer science have been identified :

1. Expanders and combinatorial optimization,
2. Complex networks and the Internet topology,
3. Data mining,
4. Computer vision and pattern recognition,
5. Internet search,
6. Load balancing and multiprocessor interconnection networks,
7. Anti-virus protection versus spread of knowledge,
8. Statistical databases and social networks,
9. Quantum computing,
10. Bio-informatics,
11. Coding theory,

12. Control theory.

This classification of numerous applications contains some overlapping in the classified material. For example, methods of data mining (in particular, spectral graph clustering) appear in computer vision, social networks and Internet search while several problems of combinatorial optimization are relevant for data mining (e.g., in clustering).

The approach in the second survey [2] is different from one in the companion paper [21]. While [21] gives a survey of the areas of applications, [2] contains a description of particular topics from the theory of graph spectra independently of the areas of Computer science in which they are used. However, for each described spectral technique fields where they are used are indicated. From the presented material one can see that a great part of the theory of graph spectra is really used in computer sciences.

The paper [2] contains, among others, the following sections and subsections:

3. Significant eigenvalues, 3.1. Largest eigenvalue, 3.2. Algebraic connectivity, 3.3. The second largest eigenvalue, 3.4. The least eigenvalue, 3.5. Main eigenvalues,

4. Eigenvector techniques, 4.1. Principal eigenvector, 4.2. The Fiedler eigenvector, 4.3. Other eigenvectors,

5. Spectral recognition problems, 6. Spectra of random graphs, 7. Miscellaneous topics, 7.1. The Hoffman polynomial, 7.2. Integral graphs, 7.3. Graph divisors.

Through various applications in computer sciences it becomes clear that spectral graph theory is by no means bounded to a particular graph matrix, such as adjacency matrix or Laplacian. A great variety of graph matrices are used depending on the problem treated.

Of course, we do not have space here to provide standard details from the theory of graphs spectra; instead we direct the reader to the corresponding mathematical literature, in particular to books [12, 17].

Since methods of Computer science are present in all branches of science, applications of graph spectral techniques to Computer science are transferred to almost all branches of science (telecommunications, electrical engineering, biology, chemistry, geography, social sciences, etc.). Sometimes by using the adjective "computational" one can denote those parts of particular sciences which overlap with Computer science (e.g., computational biology, computational chemistry, etc.). In this sense one can speak of computer sciences which explains the title of [2].

Of course, graph spectra appear in Computer science since graphs for themselves are relevant. The main benefit of using graph spectra comes from the fact that eigenvalues and eigenvectors of several graph matrices can be quickly computed (computational complexity is $O(n^3)$ where n is the number of vertices). However, spectral graph parameters contain a lot of information on the graph structure (both global and local). This includes some information on graph parameters that, in general, are computed by exponential algorithms (e.g. chromatic number, the size of maximal clique, etc.). For example, computing the chromatic number of a graph with a few thousands vertices is a difficult task while eigenvalues and eigenvectors can be computed in a few seconds (by iterative algorithms).

Graphs that are treated in computer sciences using graph spectra typically represent either some physical networks (computer network, Internet, biological network, etc.) or data structures (documents in a database, indexing structure, etc.). In the first case the graphs usually have a great number of vertices (thousands or millions) and they are called *complex networks* while in the second case graphs are of small dimensions.

This paper elaborates spectral recognition problems by extending considerations of Section 5 of the paper [2].

1.2. Preliminaries: Spectral recognition problems

The whole spectral graph theory is related in some sense to the recognition of graphs since spectral graph parameters contain a lot of information on the graph structure (both global and local). However, we shall treat here the problems of recognizing entire graphs, or some parts of them, both in an exact manner and in an approximative way.

In particular, we shall consider

- characterizations of graphs with a given spectrum
- exact or approximate constructions of graphs with a given spectrum,
- similarity of graphs,
- perturbations of graphs.

The rest of the paper is organized as follows. Section 2 discusses some differences concerning graph spectra in mathematics and computer science while in Section 3 spectral characterizations of graphs are presented. Section 4 introduces the concept of the similarity of graphs while Section 5 describes the use of the interlacing theorem. Section 6 contains a description of two kinds of graph perturbations. Some specific spectral recognition problems are described in Section 7.

2. GRAPH SPECTRA IN MATHEMATICS AND COMPUTER SCIENCE

Spectral graph theory is a very well developed mathematical field (see, for example, [17]) but also an engineering discipline [44][2].

For decades graph theory was just a collection of weakly interrelated subtheories (chromatic graph theory, metrical problems, trees, planar graphs, etc.). The theory of graph spectra contains tools which can be applied to all these subtheories, although with varying strength, and one can think of it as being a unifying theory for the whole graph theory. However, spectral techniques are weak for some problems and mathematicians could reasonably hold doubt in such a possible conclusion.

In applications to computer sciences spectral graph theory is considered as very strong and perhaps one can say that its unifying mission for graph theory has been realized through Computer Science.

---

[2] An Internet search has shown that, for example, the following universities in their computer science departments offer to students a course in spectral graph theory: Univ. Pennsylvania, Philadelphia; Yale University; Carnegy Melon University; Univ. Alicante, Spain; Duke University; Univ. Polit. Catalunia, Barcelona.

As already pointed out, the benefit of using graph spectra in treating graphs is that eigenvalues and eigenvectors of several graph matrices can be quickly computed. Spectral graph parameters contain a lot of information on the graph structure (both global and local) including some information on graph parameters that, in general, are computed by exponential algorithms.

Moreover, in some applications in data mining graph spectra are used to encode graphs themselves (see, e.g., [23, 51]). The following examples are illustrative in this respect.

The indexing structure of objects appearing in computer vision (and in a wide range of other domains such as linguistics and computational biology) may take the form of a tree. An indexing mechanism that maps the structure of a tree into a low-dimensional vector space using graph eigenvalues is developed in [40].

In some cases researchers feel that the spectrum very well characterizes the graphs under consideration so that the spectrum is considered as a fingerprint of the corresponding network. The eigenvalues $\gamma_i ; i = 1, 2, ..., n$ of the normalized Laplacian $\hat{L}$ in non-decreasing order can be represented by points $(\frac{i-1}{n-1}, \gamma_i)$ in the region $[0,1]\times[0,2]$ and can be approximated by a continuous curve. It was noticed in [45, 46] that this curve is practically the same during the time for several networks in spite of the increasing number of vertices and edges of the corresponding graph.

3. SPECTRAL CHARACTERIZATIONS OF GRAPHS

At some time it was conjectured that non-isomorphic graphs have different spectra, i.e. that graphs are characterized by their spectra. Very quickly this conjecture was refuted and numerous examples and families of non-isomorphic graphs with the same spectrum were found. In particular, it was proved that almost all trees are not characterized by their spectra. Analogous question for general graphs remained open (see, e.g., [12], Section 6.1, for a survey on these questions). Also in Chemistry there was a criticism on using graph eigenvalues to characterize molecules [30].

Graphs with the same spectrum of an associated matrix M are called *cospectral* graphs with respect to *M*, or *M* - *cospectral* graphs.

The existence of cospectral graphs is not considered as a disadvantage in using graph spectra in Computer Science since it is believed that graph spectra contain enough information for the purposes for which they are used.

To clarify recent developments we need some definitions.

A graph H cospectral with a graph G, but not isomorphic to G, is called a cospectral mate of G. Let $\mathcal{G}$ be a finite set of graphs, and let $\mathcal{G}'$ be the set of graphs in $\mathcal{G}$ which have a cospectral mate in $\mathcal{G}$ with respect to M. The ratio $|\mathcal{G}'| / |\mathcal{G}'|$ is called the *spectral uncertainty* of (graphs from) $\mathcal{G}$ with respect to M (or, in general, *spectral uncertainty* of the *M* -theory).

The papers [22, 29] provide spectral uncertainties r_n with respect to the adjacency matrix A, s_n with respect to the Laplacian L and q_n with respect to the signless Laplacian Q of sets of all graphs on n vertices for $n \le 11$ (see [5] for $n = 12$):

n	4	5	6	7	8	9	10	11	12
r_n	0	0.059	0.064	0.105	0.139	0.186	0.213	0.211	0.188
s_n	0	0	0.026	0.125	0.143	0.155	0.118	0.090	0.060
q_n	0.182	0.118	0.103	0.098	0.097	0.069	0.053	0.038	0.027

We see that the sequences s_n and q_n are decreasing for $n \le 12$ while the sequence r_n is increasing for $n \le 10$. Yet, it starts to decrease for $n > 10$. This is a strong basis for believing that almost all graphs are determined by their spectra when n tends towards the infinity, as conjectured in [22, 29].

The proof of this conjecture would strengthen the theory of graph spectra and, in particular, its application to computer sciences.

Having in view the above data, in applications the L-spectrum is used to encode graphs rather than A-spectrum, i.e. the L-spectrum has more representational power than the A-spectrum, in terms of resulting in fewer cospectral graphs. The above data show that it is even better to use signless Laplacian eigenvalues since they have stronger characterization properties.

Recently, a spectral theory of graphs based on the signless Laplacian has been developed [18, 19, 20].

There are many results in the mathematical literature on spectral characterizations of particular classes of graphs (see, e.g., Chapter 4 of [17]). For example, complete graphs, paths and circuits are characterized by their A-spectra up to an isomorphism[3]. There are also characterizations with some exceptional cospectral mates.

However, these results hardly could be applied to graphs which appear in applications to computer science.

4. SIMILARITY OF GRAPHS

There is a need to introduce the notion of similarity of graphs. (This has nothing to do with similarity in matrices, i.e. two graphs can be similar without corresponding matrices being similar.) This will be done using various distances between graphs. A special role play spectral graph distances.

Another spectrally based measure of similarity between networks has been introduced in [25], and applied to Internet topology analysis.

4.1. Spectral distances in graphs

The Euclidean distance between the eigenvalue sequences of two graphs on the same number of vertices is called the *spectral distance* of graphs. Some other spectral distances have been considered as well (e.g., the Manhattan distance).

In defining spectral distances various graph matrices can be used (e.g., the adjacency matrix, the Laplacian, the signless Laplacian).

---

[3] Graphs characterized by their spectra up to an isomorphism are called in [22, 29] DS-graphs or it is said that such graphs are DS (spectrally determined).

Some mathematical results on the Manhattan spectral distance have been obtained in [31]. An interesting observation from this paper is that the Manhattan distance is in connection with graph energy, a graph invariant studied very much in the literature. The *energy* of a graph is the sum of absolute values of its A-eigenvalues. Thus the energy of a graph is the Manhattan spectral distance of the graph from the graph without edges.

The use of the Laplacian and the signless Laplacian matrix for the Manhattan distance seems to be very useful when considering subgraphs. By the interlacing theorems for these matrices (see Section 5), all eigenvalues go down or remain the same if an edge is deleted from the graph. Hence the distance between a graph and any of its edge deleted subgraphs is equal to the decrement of the trace of the matrix. Since for both matrices the trace is equal to the sum of vertex degrees, we conclude that the distance is equal to the twofold number of deleted edges. All these properties do not hold for the adjacency matrix.

Two graphs are considered as *similar* if their spectral distance is small. If two graphs are at zero distance, this does not necessarily mean that they are equal (i.e. isomorphic); they are only cospectral. In this sense, cospectral graphs are similar. See next subsection for a possibility to introduce a positive distance between cospectral graphs.

For several reasons it is of interest to construct or generate a graph with the given spectrum.

An algorithm for such a spectral graph reconstruction has been developed in [7]. Given the spectrum of a graph, the algorithm starts from a random graph and uses the tabu search to diminish the Euclidean spectral distance between the given and the current spectrum. Both, the distance and the meta-heuristic, can be varied. We could use the Manhattan distance based on the adjacency matrix or on the signless Laplacian. The tabu search could be replaced by the variable neighbourhood search (see, for example, [6]) or by some other meta-heuristics.

The variable neighbourhood search is exploited in the programming package AutoGraphiX (briefly AGX) for finding graphs with extremal values of a graph invariant chosen by the user. The system starts from a random graph or from a graph given by the user. The graph is perturbed to some extent using the variable neighbourhood search and a new graph is chosen which improves maximally the considered graph invariant. The system AGX is very useful in formulating some conjectures which are later treated by theoretical means. For example, system AGX has generated several conjectures for the energy of a graph [6] and thirty conjectures concerning signess Laplacian eigenvalues [16]. See also [1]. It would be interesting to treat some conjectures from [31] concerning spectral distances of graphs by AGX.

AGX could be used for spectral reconstruction of graphs. It is sufficient to require that the system minimizes the distance (of any kind) between the current graph and a fixed graph. One could compare the speed of convergence for several distances and for several meta-heuristics.

Computer programs for spectral reconstruction of graphs can be used to generate example graphs with desired spectral properties.

4.2. Refinements using graph angles

Cospectral graphs are at spectral distance 0 and if we wish to define some kind of positive distance among them we need to consider graph invariants other than eigenvalues.

Since eigenvectors are not graph invariants it is reasonable to extend eigenvalue based techniques by some invariants of the eigenspaces called *graph angles*.

Let G be a graph on n vertices with distinct eigenvalues $\mu_1, \mu_2 ..., \mu_m (\mu_1 > \mu_2 > \cdots > \mu_m)$ and let $S_1, S_2, ..., S_m$ be the corresponding eigenspaces. Let $\{e_1, e_2, ..., e_n\}$ be the standard (orthonormal) basis of R^n. The numbers $\alpha_{pq} = \cos \beta_{pq} (p = 1, 2, ..., m; q = 1, 2, ..., n)$, where β_{pq} is the angle between S_p and e_q, are called *graph angles*. The sequence $\alpha_{pq} (q = 1, 2, ..., n)$ is called the *eigenvalue angle sequence* corresponding to the eigenvalue $\mu_p (p = 1, 2, ..., m)$. We also define the *angle matrix* of G, i.e. an $m \times n$ matrix (m is the number of its distinct eigenvalues, while n is the order of G) as a matrix (α_{ij}). This matrix is a graph invariant if its columns are ordered lexicographically. The rows of the angle matrix are called the *standard eigenvalue angle sequences*.

Let $x_i = (x_{i1}, x_{i2}, ..., x_{in})$ $(i = 1, 2, ..., n)$ be orthonormal eigenvectors of G. Define $M_p = \left\{ j \,\middle|\, Ax_j = \mu_p x_j \right\}$. We have $\alpha_{pq}^2 = \sum_{j \in M_p} x_{jq}^2$ for squares of angles of G. This formula holds for any choice of orthonormal eigenvectors of G (cf. [15], p. 76).

Angles between the vector $(1, 1, ..., 1) \in R^n$ and eigenspaces $S_1, S_2, ..., S_m$ are called *main angles* of the graph.

An overview of results on graph angles is given in [15] including the characterizing properties of graph angles.

It was suggested in [10] that cospectral graphs can be ordered by graph angles, in particular, lexicographically by their standard eigenvalue angle sequences. The paper provides an example of 21 mutually cospectral graphs (on 10 vertices with 20 edges) ordered by the first standard eigenvalue angle sequences.

In defining a spectral graph distance we use differences between corresponding eigenvalues of two graphs. For each spectral graph distance we can define a corresponding *cospectral graph distance* by using differences between the corresponding entries of the angle matrix instead of differences between corresponding eigenvalues. For example, the *Manhattan cospectral graph distance* is the sum of absolute values of differences between the corresponding entries of the angle matrices of the graphs.

However, the cospectral graph distance can be 0 for non-isomorphic graphs since there are graphs having the same eigenvalues and the same angles (for example, strongly regular graphs, see [15]). Such effects cannot appear if we use stronger invariants such as *canonical star basis* of a graph [9], [15]. However, the computational complexity of constructing a canonical star basis is probably so high that it is not practical to use this approach in applications.

4.3. Network alignment problem

The notion of graph similarity can be extended to graphs having different numbers of vertices. Detecting similarities between networks is frequently called *alignment* of networks. The general idea is the following one.

Given the graphs G and H, a measure of similarity between vertices of G and vertices of H is introduced by some definition. Let $R_{i,j}$ be the measure of similarity between vertex i of G and vertex j of H. Let B be the bipartite graph on vertex sets of the graphs G and H with edge weights $R_{i,j}$. We want to find a maximal matching with a maximal sum of weights in B. This matching defines subgraphs of graphs G and H which are similar w.r.t. the introduced similarity between vertices of G and vertices of H. Finding a maximal matching with a maximal sum of weights in a bipartite graph can efficiently be performed by existing algorithms for the assignment problem in combinatorial optimization (see, for example, [36]).

In this way we can find common or similar subgraphs in the considered graphs. Note that the subgraph isomorphism problem is NP-complete which means that we cannot expect a satisficatory algorithm by comparing subgraphs directly.

The measure of similarity $R_{i,j}$ is always defined taking into the account the neighbouhoods of vertex i in G and vertex j in H. For example, vertices of the same degree should be considered as being more similar than those with different degrees.

A survey of structural (non-spectral) measures of similarity between vertices can be found in [49]. The normalized positive eigenvector belonging to the largest A-eigenvalue of a connected graph is called the *principal eigenvector*.

A spectrally based measure of similarity between vertices of two graphs is used in [41] in the context of protein interaction networks. The algorithm is called IsoRank, it uses graph eigenvectors and is similar to the algorithm PageRank, used in the Internet search [4]. Here the measure of similarity $R_{i,j}$ is equal to the product[4] of the i-th coordinate of the principal eigenvector of the graph G and the j-th coordinate of the principal eigenvector of the graph H. This measure has the property that for any pair (i,j) it is equal to the mean value of similarities between all pairs (p,q) where p is a neighbour of i and q is a neighbour of j. The principal eigenvector of a graph can efficiently be computed by an iterative algorithm called the *power method* (see, for example, [28]).

A similar concept of the measure of similarity has been introduced in [3], even more generally between vertices of two digraphs and applied to synonym extraction from

---

[4] The paper [41] uses graph product $G \times H$ of graphs G and H. The quantity $R_{i,j}$ is interpreted as a coordinate of the principal eigenvector of $G \times H$. However, it is well-known that the adjacency matrix of $G \times H$ is equal to the Kronecker product of adjacency matrices of graphs G and H and that the principal eigenvector of $G \times H$ is the Kronecker product of principal eigenvectors of graphs G and H (see, for example, [12], pp. 69-70). Hence, principal eigenvectors of graphs G and H can be computed separately.

a dictionary. When considering undirected graphs, we obtain the construction used in IsoRank.

Coordinates of the principal eigenvector are related to vertex neighbourhoods because they are asymptotically proportional to the number of walks of length k starting at particular vertices. The following relevant theorem of T.H. Wei [47] is noted in [15], p. 26:

Theorem 4.1. *Let $N_k(i)$ be the number of walks of length k starting at vertex i of a non-bipartite connected graph G with vertices $1,2,...,n$. Let $s_k(i) = N_k(i) \cdot \left(\sum_{j=1}^{n} N_k(j) \right)^{-1}$. Then, for $k \to \infty$, the vector $(s_k(1), s_k(2), ..., s_k(n))^T$ tends towards the principal eigenvector of G.*

Counting walks with specified properties in a graph (or digraph) is related to graph spectra by the following well-known result (see [12], p. 44).

Theorem 4.2. *If A is the adjacency matrix of a graph, then the (i,j)-entry $a_{ij}^{(k)}$ of the matrix A^k is equal to the number of walks of length k that originate at vertex i and terminate at vertex j.*

The paper [33] surveys methods of network alignment used in protein interaction networks and recommends an algorithm based on "graphlets" (induced subgraphs on at most 5 vertices). This is again a non-spectral approach in which a vertex is characterized by a 73-dimensional vector (vertex signature) whose coordinates represent frequencies of graphlets in which the vertex appears. The corresponding programming package is called GRAAL.

Let G be a graph with adjacency matrix A, and let $N_k(j) = a_{jj}^{(k)}$ the number of walks of length k in G originating and terminating at vertex j. Let $\mathcal{H}_j(t)$ be the generating function $\sum_{k=0}^{\infty} N_k(j) t^k$. Using notation from Subsection 4.2, we can obtain, [15], p. 82,

$$\mathcal{H}_j(t) = \sum_{k=0}^{\infty} t^k \sum_{i=1}^{m} \alpha_{ij}^2 \mu_i^k = \sum_{i=1}^{m} \frac{\alpha_{ij}^2}{1 - \mu_i t}.$$

On the other hand, we have $\mathcal{H}_j(t) = 1 + d_j t^2 + 2t_j t^3 + ...,$ where d_j is the degree of vertex j and t_j is the number of triangles containing j. The quantity t_j is also called the *clustering coefficient* of the vertex j. Higher terms of $H_j(t)$ give the numbers of closed walks contained in graphlets with 4 and 5 vertices to which the vertex j belongs.

We think that in problems of network alignment vertices should be characterized by generating functions $\mathcal{H}_j(t)$. This function depends on the vertex neighbourhood which is in this case extended to the whole graph unlike the method with graphlets where the neighbourhood is very limited.

For example, the measure of similarity $R_{i,j}$ can be defined in some way using the difference $\mathcal{H}_i^G(t) - \mathcal{H}_j^H(t)$ of generating functions of vertex i in G and vertex j in H.

The following formulas are also useful.

$$N_s(j) = a_{jj}^{(s)} = \sum_{i=1}^{n} \mu_i^s \alpha_{ij}^2.$$

The degree d_j of the vertex j, and the number t_j of triangles containing the vertex j, are given by

$$d_j = \sum_{i=1}^{m} \alpha_{ij}^2 \mu_i^2, \quad t_j = \frac{1}{2} \sum_{i=1}^{m} \alpha_{ij}^2 \mu_i^3.$$

Let $P_G(\lambda) = \det(\lambda I - A)$ be the characteristic polynomial of the graph G. The generating function can be obtained by the formula

$$\mathcal{H}_j^G(t) = P_{G-j}(\frac{1}{t}) / tP_G(\frac{1}{t}),$$

since

$$P_{G-j}(x) = P_G(x) \sum_{i=1}^{m} \frac{\alpha_{ij}^2}{x - \mu_i}.$$

It would be interesting to compare the performances of the three described approaches to the network alignment problem using various criteria. It was reported in [33] that GRAAL finds common subgraphs with more vertices than IsoRank.

It is interesting to note that the use of the signless Laplacian (Manhattan) spectral distance provides an upper bound for the number of edges in a common subgraph of two graphs. By the Interlacing theorem, the signless Laplacian eigenvalue sequence of a maximal common subgraph is dominated by signless Laplacian eigenvalue sequences of each of the considered networks. Hence, an upper bound for the number of edges in a common subgraph is equal to $\frac{1}{2}\sum \kappa_i$, where κ_i is the minimum of the i-th largest signless Laplacian eigenvalues of the considered networks.

5. QUERIES FOR DATABASES AND THE SUBGRAPH ISOMORPHISM PROBLEM

In several databases the data are often represented as graphs. Very frequently graphs are indexed by their spectra.

In [37] a spectral graph theory approach is presented for representing melodies as graphs, based on intervals between the notes they are composed of. These graphs are then indexed using their Laplacian spectrum. This makes it possible to find melodies similar to a given melody. See also [40], [51] for other examples.

The query for such a database is given by a graph. To find similar data in the database it is necessary to compare subgraphs of the query graph with subgraphs of the graphs stored in the database. One should efficiently select a small set of database graphs, which share a subgraph with the query.

The *subgraph isomorphism problem* is a computational task in which two graphs G and H are given as input, and one must determine whether G contains a subgraph that is isomorphic to H. Sometimes the name *subgraph matching* is also used for the same problem. Subgraph isomorphism is a generalization of two well-known NP-complete problems, the maximum clique problem and the problem of testing whether a graph contains a Hamiltonian cycle, and is therefore NP-complete [8]. There is a similar problem known as the *maximum common subgraph-isomorphism* problem. This problem is known to be NP-hard. The formal description of the problem is as follows: given two graphs G_1 and G_2, what is the largest induced subgraph of G_1 isomorphic to an induced subgraph of G_2? The associated decision problem, i.e., given G_1, G_2 and an integer k, deciding whether G_1 contains an induced subgraph of at least k edges isomorphic to an induced subgraph of G_2 is NP-complete.

Instead of comparing subgraphs one can compare their spectra. While the subgraph isomorphism problem is NP-complete, comparing spectra can be done in polynomial time. The so called Interlacing Theorem (Theorem 5.1) plays an important role in problems of spectral graph recognition and in spectral graph theory and its applications in general.

Recall that the matrix A with complex entries a_{ij} is called *Hermitian* if $A^T = \overline{A}$, i.e. $a_{j,i} = \overline{a}_{ij}$ for all *i,j*.

Theorem 5.1. *(see, e.g., [12], p. 19) Let A be a Hermitian matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and let B be one of its principal submatrices. If the eigenvalues of B are $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m$ then $\lambda_{n-m+1} \leq \mu_i \leq \lambda_i (i = 1,...,m)$.*

The inequalities of this theorem are known as *Cauchy's inequalities* and the whole theorem is known as the *Interlacing Theorem*.

Usually, A is the adjacency matrix of a graph G and B is the adjacency matrix of an induced subgraph H of the graph G.

We have the following version of the interlacing theorem for L- spectra.

Theorem 5.2. *Let G be a connected graph on n vertices. Eigenvalues in non-decreasing order of the Laplacian $L = D - A$ of G are denoted by $v_1 = 0, v_2, ..., v_n$. Let G' be obtained from G by adding an edge and let $\sigma_1 = 0, \sigma_2, ..., \sigma_n$ be L-eigenvalues of G'. Then*

$$0 = v_1 = \sigma_1 \leq v_2 \leq \sigma_2 \leq \cdots \leq v_n \leq \sigma_n.$$

The proof is obtained using well-known Courant-Weyl inequalities (see, e.g., [12], pp. 51-52).

Literally the same theorem holds for the signless Laplacian spectra but the proof relies on some connections between Q-eigenvalues and A-eigenvalues of line graphs [16].

The interlacing theorem is often an effective tool in pruning the search in graph databases. Such databases consist of graphs and a query is also a graph Q. One should find all graphs in the database which contain the graph Q as a subgraph.

The problem is that Cauchy's inequalities hold for induced subgraphs and not necessarily for subgraphs in general. However, in the set of trees connected subgraphs are always induced subgraphs and the pruning rules based on the Interlacing theorem do work [40].

Using this fact a spectral coding of graphs is introduced in [51]. This coding uses spectra of trees associated to graph vertices. It seems that better search pruning possibilities can be expected from interlacing theorems for Laplacian and signless Laplacian spectra. Laplacian eigenvalues are used to code graphs, for example, in [37], [23].

To accelerate the process of computing spectra of subgraphs the *spectral integral variation technique* is used in [23].

By the interlacing theorem, when adding an edge to a graph the L-eigenvalues do not decrease. However, the sum of L-eigenvalues increases by 2. We are interested in the case when L-eigenvalues change only by integer quantities. Evidently, there are just two possible scenarios [43, 24] where that can happen: either one eigenvalue will increase by 2 (and $n-1$ eigenvalues remain unchanged) or two eigenvalues will increase by 1 (and $n-2$ eigenvalues remain unchanged). Precise conditions when each of these two cases of spectral integral variation technique occurs are given in the literature [32, 43].

The use of the signless Laplacian eigenvalues looks even better.

6. STRUCTURAL AND SPECTRAL PERTURBATIONS OF GRAPHS

Informally speaking, similar graphs can be understood as being obtained, one from the other, by some"small" perturbation. A graph perturbation can be described as a "small" change in its structure or in its spectrum.

The spectral integral variation technique, described in the last section, is just an example involving *graph perturbations*. A graph perturbation means a small change in graph structure (e.g., adding an edge or a vertex). We are interested in changes in graph eigenvalues caused by a perturbation.

There is a chapter in the book [15] devoted to graph perturbations and corresponding changes in the spectrum.

The problem of protecting the privacy appears in *social networks* on the Internet (for example, Facebook) when studying general properties of an existing network. A way to protect the privacy of personal data is to randomize the network representing relations between individuals by deleting some actual edges and by adding some additional edges in such a way that the global characteristics of the network are unchanged. This is achieved by using eigenvalues of the adjacency matrix (in particular, the largest one) and of the Laplacian (algebraic connectivity) to control the process of deleting and adding the edges [50]. The choice of deleted and added edges is performed by using results of [15], Chapter 6, for the largest eigenvalue and the corresponding results for the algebraic connectivity have been derived in the paper.

In Computer Science literature, some spectral perturbations of graphs have been considered as well. This means that the graph spectrum is slightly changed while the eigenvectors remain unchanged. This is used in connection with the formula for spectral decomposition of the adjacency matrix A of a graph, i.e. $A = U\Lambda U^T$; where Λ is a

diagonal matrix containing the eigenvalues of A and the columns of matrix U are orthonormal eigenvectors of A. The paper [34] proposes a new robustness parameter for complex networks: this is the maximal number k such that one can replace k smallest in modulus eigenvalues of A with zeros with the possibility that A still can be reconstructed.

A similar "deletion" of eigenvalues appears also in the so called *latent semantic indexing* (LSI) but it is applied on singular values of the term-by-document matrix (see, e.g., [35, 42]).

7. SOME OTHER SPECTRAL RECOGNITION PROBLEMS

The second smallest Laplacian eigenvalue is called *algebraic connectivity* of the graph and was introduced by Fiedler [26]. The eigenvector belonging to the second smallest Laplacian eigenvalue of the connected graph is called the *Fiedler eigenvector*. Of course, we have both positive and negative entries in it.

A heuristic for solving the min-cut problem uses the Fiedler eigenvector to partition the vertex set into parts corresponding to positive and negative coordinates of this vector [27].

These ideas were exploited in the literature in various ways for devising powerful heuristics for spectral graph partitioning and/or clustering. For instance, Shi and Malik [39] have shown how the sign pattern of the Fiedler eigenvector can be used to separate the foreground from the background structure in images. The original procedure from [27] has been improved by using the matrix $D^{-1}L$ (so as to maximize the normalized graph cut). More generally, image segmentation is an important procedure in computer vision and pattern recognition. The problem is to divide the image into regions according to some criteria. Very frequently the image segmentation is obtained using eigenvectors of some graph matrices (for more details see, e.g., [48]).

Graph spectra can be used to recognize some abstract objects like hard instances for a combinatorial optimization problem.

Let A be an (exact) algorithm for solving an NP-hard combinatorial optimization problem C and let I be an instance of C of dimension n. A complexity index of I for C with respect to A is a real r, computable in polynomial time from I, by which we can predict (in a well defined statistical sense) the execution time of A for I.

We consider the symmetric travelling salesman problem with instances I represented by complete graphs G with distances between vertices (cities) as edge weights (lengths). Intuitively, the hardness of an instance G depends on the distribution of short edges within G. Therefore we consider some short edge subgraphs of G (minimal spanning tree, critical connected subgraph, critical 2-connected subgraph and several others) as non-weighted graphs and several their invariants as potential complexity indices. Here spectral invariants (e.g. spectral radius of the adjacency matrix) play an important role since, in general, there are intimate relations between eigenvalues and the structure of a graph. Since hidden details of short edge subgraphs really determine the hardness of the instance, one should use techniques of data mining to find them. In particular, spectral clustering algorithms are used including information obtained from the spectral gap in Laplacian spectra of short edge subgraphs [11].

Acknowledgement. This work is supported by the Serbian Ministry for Education and Science, Grants ON174033 and III44006.

REFERENCES

[1] Aouchiche M., and Hansen P., "A survey of automated conjectures in spectral graph theory", *Linear Algebra Appl.*, 432(2010) 2293-2322.
[2] Arsić B., Cvetković D., Simić S.K., and Škarić M., "Graph spectral techniques in computer sciences", *Applicable Analysis and Discrete Mathematics*, 6 (1)(2012) 1-30.
[3] Blondel V.D., Gajardo A., Heymans M., Senellart P., and Van Doren P., "A measure of similarity between graph vertices: applications to synonym extraction and web searching", *SIAM Rev*., 46(4)(2004) 647-666.
[4] Brin S., and Page L., "The Anatomy of Large-Scale Hypertextual Web Search Engine", Proc. 7th International WWW Conference, 1998.
[5] Brouwer A.E., and Spence E., "Cospectral graphs on 12 vertices", *Electron. J. Combin*., 16, 2009.
[6] Caporossi G., Cvetković D., Gutman I., and Hansen P., "Variable neighborhood search for extremal graphs, 2. Finding graphs with extremal energy", *J. Chem. Inform. Comp. Sci.,* 39(1999) 984-996.
[7] Comellas F., and Diaz-Lopez J., "Spectral reconstruction of complex networks", *Phys. A*, 387(2008) 6436-6442.
[8] Cook, S. A. , "The complexity of theorem-proving procedures", *Proc. 3rd ACM Symposium on Theory of Computing*, 1971, 151-158.
[9] Cvetković D., "Star partitions and the graph isomorphism problem", *Linear and Multilinear Algebra*, 39(1-2)(1995) 109-132.
[10] Cvetković D., "Characterizing properties of some graph invariants related to electron charges in the Hüuckel molecular orbital theory", *Proc. DIMACS Workshop on Discrete Mathematical Chemistry*, *DIMACS Ser. Discrete Math. Theoret. Comp, Sci.*, 51(2000) 79-84.
[11] Cvetković D., "Complexity indices for the travelling salesman problem and data mining", *Transactions of Combinatorics*, 1(1)(2012) 35-43.
[12] Cvetković D., Doob M., and Sachs H., *Spectra of Graphs, Theory and Application*, 3rd edition, Johann Ambrosius Barth Verlag, Heidelberg-Leipzig, 1995.
[13] Cvetković D., and Gutman I., (eds.), *Applications of Graph Spectra*, *Zb. Rad. (Beogr.)*, 13(21) Mathematical Institute SANU, Belgrade, 2009.
[14] Cvetković D., and Gutman I., (eds.), *Selected Topics on Applications of Graph Spectra*, *Zb. Rad. (Beogr.)*, 14(22), Mathematical Institute SANU, Belgrade, 2011.
[15] Cvetković D., Rowlinson P., and Simić S. K., *Eigenspaces of Graphs*, Cambridge University Press, Cambridge, 1997.
[16] Cvetković D., Rowlinson P., and Simić S., "Eigenvalue bounds for the signless Laplacian", *Publ. Inst. Math.* (Beograd) 81(95)(2007) 11-27.
[17] Cvetković D., Rowlinson P., and Simić S. K., *An Introduction to the Theory of Graph Spectra*, Cambridge University Press, Cambridge, 2009.
[18] Cvetković D., and Simić S.K., "Towards a spectral theory of graphs based on the signless Laplacian I", *Publ. Inst. Math.* (Beograd) 85(99)(2009) 19-33.
[19] Cvetković D., and Simić S.K., "Towards a spectral theory of graphs based on the signless Laplacian, II", *Linear Algebra Appl.*, 432(2010) 2257-2272.
[20] Cvetković D., and Simić S.K., "Towards a spectral theory of graphs based on the signless Laplacian, III", *Appl. Anal. Discrete Math.*, 4(2010) 156-166.
[21] Cvetković D., and Simić S.K., "Graph spectra in computer science", *Linear Algebra Appl.*, 434(2011) 1545-1562.

[22] Dam E.R. van, and Haemers W., "Which graphs are determined by their spectrum? ", *Linear Algebra Appl.*, 373(2003) 241-272.
[23] Demirci M.F., Leuken R.H. van, and Veltkamp R.C., "Indexing through laplacian spectra", *Computer Vision and Image Understanding*, 2008. doi: 10.1016/j.cviu.2007.09.012.
[24] Fan Y., "On spectral integral variations of graphs", *Linear Multilinear Algebra* 50 (2002) 133-142.
[25] Fay D., Haddadi H., Thomason A., Mortier R., Jamakovic A., Uhlig S., and Rio M., "Weighted spectral distribution for internet topology analysis: theory and applications", *J. IEEE/ACM Trans. Networking (TON)*, 18(1)(2010) 164-176.
[26] Fiedler M., "Algebraic connectivity of graphs", *Czechoslovak Math. J.*, 23(98)(1973) 298-305.
[27] Fiedler M., "A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory", *Czechoslovak Math. J.*, 25(100)(1975) 619-633.
[28] Golub G.H., and Van Loan C., *Matrix Computations*, Johns Hopkins University Press, 2006.
[29] Haemers W., and Spence E., "Enumeration of cospectral graphs", *European J. Combin.* 25 (2004) 199-211.
[30] Heilbronner E., and Jones T.B., "Spectral differences between "isospectral" molecules", *J. Amer. Chem. Soc.*, 100(20)(1978) 6505-6507.
[31] Jovanović I., and Stanić Z., "Spectral distances of graphs", *Linear Algebra Appl.*, 436(2012) 1425-1435.
[32] Kirkland S., "A Characterization of Spectral Integral Variation in Two Places for Laplacian Matrices", *Linear Multilinear Algebra*, 52 (2004) 7998.
[33] Kuchaiev O., Milenković T., Memišević V., Hayes W., and Pržulj N., "Topological network alignment uncovers biological function and phylogeny", *J. Roy. Soc. Interface*, 7(2010) 1341-1354.
[34] Liu D., Wang H., and Van Mieghem P., "Spectral perturbation and reconstructability of complex networks", *Phys. Rev. E*, 81(2010) 016101, 1-9.
[35] Papadimitriou, C.H., Raghavan P., Tamaki H., and Vempala S., Liu D., Wang H., and Van Mieghem P., "Spectral perturbation and reconstructability of complex networks"Latent semantic indexing: a probabilistic analysis Liu D., Wang H., and Van Mieghem P., "Spectral perturbation and reconstructability of complex networks", *J. Comput. System Sci.*, 61(2)(2000) 217-235.
[36] Papadimitriou, C., and Steiglitz K., *Combinatorial Optimization: algorithms and complexity*, Dover, 1998.
[37] Pinto, A., MIREX2007 - *Graph spectral method*, to appear.
[38] Pržulj, N., Corneil, D.G., and Jurisica, I., "Efficient estimation of graphlet frequency distributions in proteinprotein interaction networks", *Bioinformatics*, 22(8)(2006) 974980.
[39] Shi, J., and Malik, J., "Normalized cuts and image segmentation", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1997, 731-737; *IEEE Trans. Pattern Analysis Machine Intell.*, 28(2000) 888-905.
[40] Shokoufandeh, A., Dickinson, S. J., Siddiqi, K., and Zucker, S. W., "Indexing using a spectral encoding of topological structure", *IEEE Trans. Comput. Vision Pattern Recognition*, 2 (1999) 491-497.
[41] Singh, R., Xu J., and Berger B., "Pairwise global alignment of protein interaction networks by matching neighborhood topology", *Research in Computational Molecular Biology*, Springer, 2007, 16-31.
[42] Skilicorn, D.B., *Understanding Complex Data Bases: Data Mining with Matrix Decompositions, Ch*apman & Hall/CRC, New York, 2007.
[43] So, W., "Rank one perturbation and its application to the Laplacian spectrum of a graph", *Linear Multilinear Algebra*, 46 (1999) 193 198.
[44] Spielman, D. A., "Spectral graph theory and its applications", *48th Annual IEEE Symposium on Foundations of Computer Science, IEEE*, 2007, 29-38.

[45] Vukadinović, D., Huang, P., and Erlebach, T., "A spectral analysis of the Internet topology", Technical Report 118, TIK Institute, ETH Zürich, 2001.
[46] Vukadinović, D., Huang P., and Erlebach T., "On the spectrum and structure of the Internet topology graphs", *Proc. Second Internat. Workshop on Innovative Internet Computing Systems, IICS '02*, 2346 (2002) 83-95.
[47] Wei, T.H., "The algebraic foundations of ranking theory", Thesis, Cambridge, 1952.
[48] Weiss, Y., "Segmentation using eigenvectors: a unifying view", *Proc. Seventh IEEE Internat. Conf. Computer Vision*, 2 (1999) 975-982.
[49] Xuan, Q., Yu, L., Du F., and Wu, T.-J., "*A* review on node-matching between networks", in: Y. Zhang, (ed.), *New Frontiers in Graph Theory*, Intex, Rijeka, 2012, 153-167.
[50] Ying, X., and Wu, X., "Randomizing social networks: a spectrum preserving approach", *Proc. SIAM Internat. Conf. Data Mining, SDM2008*, April 24{26, 2008, Atlanta, Georgia, USA, SIAM, 2008, 739-750.
[51] Zou, L., Chen, L., Yu, J.X., and Lu, Y., "A novel spectral coding in a large graph database", *EDBT'08*, March 25 - 30, 2008, Nantes, France.