Linear Algebra and its Applications 434 (2011) 1545–1562

Contents lists available at ScienceDirect

Linear Algebra and its Applications

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

Graph spectra in Computer Science ☆

Dragoš Cvetković *, Slobodan Simić

Mathematical Institute SANU, P.O. Box 367, 11000 Belgrade, Serbia

ARTICLE INFO

Article history:
Received 6 July 2010
Accepted 18 November 2010
Available online 24 December 2010

Submitted by R.A. Brualdi

AMS classification:
05C50
15A18
15A51
65F15
68M07
68P20
68R10
68U10
81P68

Keywords:
Graph theory
Graph spectra
Applications
Computer Science
Information technology
Communication technology
Internet
Complex networks

ABSTRACT

In this paper, we shall give a survey of applications of the theory of graph spectra to Computer Science. Eigenvalues and eigenvectors of several graph matrices appear in numerous papers on various subjects relevant to information and communication technologies. In particular, we survey applications in modeling and searching Internet, in computer vision, data mining, multiprocessor systems, statistical databases, and in several other areas. Some related new mathematical results are included together with several comments on perspectives for future research. In particular, we claim that balanced subdivisions of cubic graphs are good models for virus resistent computer networks and point out some advantages in using integral graphs as multiprocessor interconnection networks.

☆ Supported by the Serbian Ministry for Science and Technological Development, Grant 144015G.

* Corresponding author.
E-mail addresses: ecvetkod@kondor.etf.rs (D. Cvetković), sksimic@mi.sanu.ac.rs (S. Simić).

0024-3795/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2010.11.035

1. Introduction

In this paper, we shall give a survey of applications of the theory of graph spectra to Computer Science.

Note that we are not giving a survey on applications of matrices in Computer Science, or on applications of graphs in Computer Science. We want to survey applications of the *theory of graph spectra* (or of *spectral graph theory*) in Computer Science.

Spectral graph theory is a mathematical theory where linear algebra and graph theory meet together.

A spectral graph theory is a theory in which graphs are studied by means of eigenvalues of a matrix M which is in a prescribed way defined for any graph. This theory is called *M-theory*.

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. Some other graph matrices will be introduced later.

The spectral graph theory includes all these particular theories together with interaction tools.

It was recognized in about last 10 years that graph spectra have several important applications in Computer Science. Graph spectra appear in internet technologies, pattern recognition, computer vision, data mining, multiprocessor systems, statistical databases and in many other areas. There are thousands of such papers.

Applications of graph spectra, in Computer Science and in general, are so numerous that we cannot give a comprehensive survey in limited space that we have at the disposal. We shall rather limit ourselves to review representative examples of applications so that the reader can get an impression on the situation and also become able to use the literature.

Several papers in Computer Science cite books on graph spectra such as [7,18,28,31]. To document spectral techniques used several books on matrices are cited as well.

Spectral techniques appear in many papers in Computer Science, perhaps there are several thousands of such papers. However, spectral techniques are far from being exclusive or essential in most cases; they are interlaced with other mathematical tools.

One should be noted that spectra of several graph matrices appear in applications. The adjacency matrix and Laplacian appear most frequently but also the signless Laplacian as well as normalized versions of these matrices. Incidence, distance and other matrices can be found as well. Sometimes the considerations move from graph matrices to general ones; equivalently, weighted graphs appear instead of graphs. In some cases we encounter digraphs and hyper-graphs as well.

Several models of random graphs together with the corresponding eigenvalue distributions appear in the treatment of complex networks (networks with a huge number of vertices).

It can be noticed that not only the eigenvalues but also the eigenvectors of relevant graph matrices appear in applications in most cases.

Several researchers in Computer Science declare that spectral graph theory is (one of) their scientific field(s). Note that mathematicians usually speak of the theory of graph spectra or just graph spectra.

Section 2 with several subsections surveys briefly main applications of graph spectra in Computer Science. Section 3 contains several comments including suggestions for further work and some relevant mathematical results. We present some general references on applications and give some bibliographic comments in Section 4.

2. A survey of applications

It is difficult to classify the existing applications of graph spectra to Computer Science. Our survey is organized through the following subsections:

2.1. Expanders and combinatorial optimization.
2.2. Complex networks and the Internet.
2.3. Data mining.
2.4. Computer vision and pattern recognition.

2.5. Internet search.
2.6. Load balancing and multiprocessor interconnection networks.
2.7. Anti-virus protection vs. spread of knowledge.
2.8. Statistical databases and social networks.
2.9. Quantum computing.

Note that this classification of numerous applications into subsections contains some overlapping of the classified material. For example, methods of data mining appear in computer vision and Internet search while several problems of combinatorial optimization are relevant for data mining (e.g., in graph clustering).

2.1. Expanders and combinatorial optimization

One of the oldest applications (from 1970s) of graph eigenvalues in Computer Science is related to graphs called *expanders*. Informally, we shall say that a graph has good *expanding properties* if each subset of the vertex set of small cardinality has a set of neighbors of large cardinality. Expanders and some related graphs (called *enlargers*, *magnifiers*, *concentrators* and *super-concentrators*, just to mention some specific terms) appear in treatment of several problems in Computer Science (for example, communication networks, error-correcting codes, optimizing memory space, computing functions, sorting algorithms, etc.). Expanders can be constructed from graphs with a small second largest eigenvalue in modulus. Such class of graphs includes the so called *Ramanujan graphs*.[1] Paper [63] is one of the most important papers concerning Ramanujan graphs.

For an introduction to this type of applications see [36] and references cited therein.

Expanders are related to some problems of combinatorial optimization. More generally, several algorithms of combinatorial optimization are considered as a part of Computer Science.

A *sensor network* consists of spatially distributed sensors (with limited capacities) and links connecting them. One of the basic problems with these networks is to design a topology (connection graph) that maximizes the ratio $\frac{\nu_2}{\nu_n}$, where ν_n is the largest while ν_2 the second smallest eigenvalue of the graph Laplacian. The larger is this ratio, the faster is the convergence speed of the decision fusion algorithm, and thus better the performance of the network. In [57], it was pointed that (non-bipartite) Ramanujan graphs are good candidates for desired topologies. There are many other problems in sensor networks where the tools from the combinatorial optimization and spectral graph theory can help, say in solving partitioning, assignment, routing and scheduling problems.

Numerous relations between eigenvalues of graphs and *combinatorial optimization* have been known for last 20 years. The section titles of an excellent expository article [67] show that many problems in combinatorial optimization can be treated using eigenvalues: 1. Introduction, 1.1. Matrices and eigenvalues of graphs; 2. Partition problems; 2.1 Graph bisection, 2.2. Connectivity and separation, 2.3. Isoperimetric numbers, 2.4. The maximum cut problem, 2.5. Clustering, 2.6. Graph partition; 3. Ordering, 3.1. Bandwidth and min-p-sum problems, 3.2. Cut-width, 3.3 Ranking, 3.4. Scaling, 3.5. The quadratic assignment problem; 4. Stable sets and coloring, 4.1. Chromatic number, 4.2. Lower bounds on stable sets, 4.3. Upper bounds on stable sets, 4.4. k-colorable subgraphs; 5. Routing problems, 5.1. Diameter and the mean distance, 5.2. Routing, 5.3. Random walks; 6. Embedding problems; A. Appendix: computational aspects; B. Appendix: rigenvalues of random graphs. The paper [67] contains a list of 135 references.

Note that one of early heuristics for graph bisection uses the *Fiedler vector*, i.e., the eigenvector belonging to the second smallest eigenvalue of the graph Laplacian. This eigenvalue is called *algebraic connectivity* of the graph and was introduced by Fiedler [44].

The algebraic connectivity has been used in [22] to formulate the following discrete semi-definite programming model of the symmetric *travelling salesman problem* (STSP):

[1] Let $\Lambda(G)$ be the second largest modulus of an eigenvalue of a graph G. A *Ramanujan graph* is a connected regular graph of degree r for which $\Lambda(G) \leqslant 2\sqrt{r-1}$.

$$\text{minimize } F(X) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(-\frac{1}{2}d_{ij}\right)x_{ij} + \frac{\alpha}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}d_{ij}$$

subject to

$$\begin{aligned}
&x_{ii} = 2+\alpha-\beta \quad (i=1,\ldots,n),\\
&\sum_{j=1}^{n} x_{ij}\alpha - \beta \quad (i=1,\ldots,n),\\
&x_{ij} \in \{\alpha-1,\alpha\} \ \ (j=1,\ldots,n:\ i<j),\\
&X \geqslant 0.
\end{aligned}$$

Here $X \geqslant 0$ means that the matrix X is symmetric and positive semi-definite, while α and β are chosen so that $\alpha > h_n/n$ and $0 < \beta \leqslant h_n$ with $h_n = 2-2\cos(2\pi/n)$ being the algebraic connectivity of the cycle C_n.

See [28, pp. 417–418,83] for further data and references.

2.2. Complex networks and the Internet

Complex networks is a common name for various real networks which are presented by graphs with an enormously great number of vertices. Here belong Internet graphs, phone graphs, e-mail graphs, social networks and many other. In spite of their diversity such networks share some common properties.

Several models of random graphs have been used to describe complex networks including the classical Erdös-Rényi model where we have a constant probability for the existence of each edge. There are models where given degree distribution is realized.

Main characteristic of complex networks is the degree and eigenvalue distribution. Both distributions obey a *power law* of the form $x^{-\beta}$ for a positive constant β.

In particular, if n_k denotes the number of vertices of degree k, then asymptotically $n_k = ak^{-\beta}$ for some constant a.

It was conjectured in [43] that in networks with degree power law the largest eigenvalues of the adjacency matrix have also a power law distribution. That was proved under some conditions in [66].

The power law for eigenvalues can be formulated in the following way. Let $\lambda_1, \lambda_2, \ldots$ be non-increasing sequence of eigenvalues of the adjacency matrix, then asymptotically $\lambda_k = ak^{-\gamma}$ for some constant a and positive γ.

The book [19] is devoted to complex networks. There are two chapters which describe spectral properties of such networks. The forthcoming book [87] describes how graph spectra are used in complex networks. See also [47].

Note that most of the papers on complex networks appear in scientific journals in the area of Physics.

The Internet is a collection of thousands of local networks (Autonomous Systems, AS) of computers (hosts and routers). Autonomous Systems are linked by a common set of protocols which enable communication and allow the use of services located at any of the other Autonomous Systems.

The whole Internet, or a part of it, can be represented by a graph in which the vertices correspond to hosts and routers while the edges correspond to physical connections between them. In another representations the vertices correspond to Autonomous Systems and the edges to the links.

Studying and modeling Internet topology (i.e., the structure) is necessary for protocol performance evaluation and simulation of a variety of network problems. Although real topology data are partially available (e.g., at the level of the so called Autonomous Systems) it is also useful to have theoretical models. Of course, theoretical models are checked on available real data. The main theoretical models of the Internet use the concepts of complex networks and, in particular, power laws for degrees and eigenvalues.

Empirical studies of the Internet topology have been conducted in many papers using the normalized Laplacian matrix $\hat{L} = D^{-\frac{1}{2}}(D-A)D^{-\frac{1}{2}} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$. This matrix has 1's on the diagonal, and at

an off-diagonal position (i, j) the entry is equal to 0 for non-adjacent and $-\frac{1}{\sqrt{d_i d_j}}$ for adjacent vertices i, j of degrees d_i, d_j. The spectrum of $\hat{L}$ belongs to the interval $[0, 2]$ independently of the number of vertices. The book [18] is devoted to the normalized Laplacian.

The eigenvalues $\gamma_i;\ i = 1, 2, \ldots, n$ of $\hat{L}$ in non-decreasing order can be represented by points $\left(\frac{i-1}{n-1}, \gamma_i\right)$ in the region $[0, 1] \times [0, 2]$ and can be approximated by a continuous curve. It was noticed in [89,90] 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. Therefore the authors consider the spectrum of $\hat{L}$ as a *fingerprint* of the corresponding network topology.

Two particular data sets on sections of the internet graph have been studied in [15] using eigenvalues of the normalized Laplacian matrix.

Analysis of the collected data sets indicates that the Internet topology is characterized by the presence of various power-laws observed when considering vertex degrees vs. vertex rank (vertices ordered by non-increasing degrees), vertex degree frequency vs. degree, and the number of vertices within a number of hops vs. number of hops [43,77]. Some of these early conclusions were subsequently revised by considering a more complete AS-level representation of the Internet topology [16,14]. These extended data diagrams have heavy tailed or highly variable degree distributions and only the distribution tales have the power-law property. It has been observed that the power-law exponents associated with Internet topology have not substantially changed over the years in spite of the Internet exponential growth [49,68]. Power-laws also appear in the eigenvalues of the adjacency matrix and the normalized Laplacian matrix. They also show invariance regardless of the rapid growth of the Internet.

Several other data on using graph eigenvalues in theoretical and empirical studying Internet topology can be found in the literature.

2.3. Data mining

Data mining discovers interesting and unknown relationships and patterns in huge data sets. Such hidden information could contribute very much to many domains such as image processing, web searching, computer security and many others including those outside Computer Science.

Among many tools used in data mining, spectral techniques play an important role [72,81]. Here belong, in particular, clustering and ranking the vertices of a graph. While ranking will be treated in 2.5, here we consider the clustering.

A description of *spectral clustering* methods is given in the tutorial [65].

Here we assume that the data are represented by a graph. There are several procedures to construct such a graph in the case that the data are given as points with mutual distances or with a similarity function.

We shall present an algorithm for graph clustering which is based on the Laplacian matrix of a graph. Similar clustering algorithms based on the normalized Laplacian or other graph matrices have been formulated in the literature.

Let G be a connected graph on n vertices. Eigenvalues in non-decreasing order and corresponding orthonormal eigenvectors of the Laplacian $L = D - A$ of G are denoted by $\nu_1 = 0, \nu_2, \ldots, \nu_n$ and $u_1, u_2, \ldots, u_n$, respectively.

In order to construct k clusters in a graph we form an $n \times k$ matrix U containing the vectors $u_1, u_2, \ldots, u_k$ as columns. In this way we have constructed a geometric representation $\mathcal{G}$ of G in the k-dimensional space $\mathbf{R}^k$: we just take rows of U as point coordinates representing the vertices of G. Edges are straight line segments between the corresponding points. Now classical clustering methods (say k-means algorithm) should be applied to this new graph presentation.

The results obtained by this and similar spectral clustering algorithms are very good and popular among researchers. However, these algorithms are not completely theoretically explained and understood.

We shall explain some basic facts on why such clustering algorithms perform well.

The following well-known inequality for the Rayleigh quotient

$$\nu_1 \leqslant \frac{\mathbf{x}^T L\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \leqslant \nu_n$$

holds for any non-zero vector $\mathbf{x}$ of the corresponding dimension. Equality holds for relevant eigenvectors.

More generally, the eigenvalue ν_i is the minimal value of the Rayleigh quotient of L over the orthogonal complement of the subspace generated by eigenvectors $u_1, u_2, \ldots, u_{i-1}$.

Note that if $\mathbf{x}^T = (x_1, x_2, \ldots, x_n)$, then

$$\mathbf{x}^T L\mathbf{x} = \sum_{i\sim j,\ i<j} (x_i - x_j)^2.$$

We also have

$$\nu = \sum_{i\sim j,\ i<j} (x_i - x_j)^2$$

if $\mathbf{x}$ is a normalized eigenvector belonging to eigenvalue ν of L.

It is now easy to see that the sum of squares of lengths of all edges in the representation $\mathcal{G}$ of G is equal to $\nu_1 + \nu_2 + \cdots + \nu_k$, as noted in the literature (see, for example [61]). It is clear that we have achieved a minimal value over all representations obtained via matrix U with orthonormal columns.

It should be expected that such an extremal graph representation must have remarkable properties. In particular, this representation enhance the cluster-properties of the original data and clusters can now be trivially detected.

In regular graphs we can use the adjacency matrix A instead of the Laplacian L. Namely, we have $L = rI - A$ for a regular graph of degree r and $\lambda_i = r - \nu_i$ for $i = 1, 2, \ldots, n$. Eigenvectors of L are also eigenvectors of A for the corresponding eigenvalues. Instead of first k smallest eigenvalues we have to consider now k largest eigenvalues.

Example. Consider the adjacency matrix A of a cycle of C_n length n. It is well known that the eigenvalues λ_j of the matrix A are given by $\lambda_j = 2\cos(2\pi j/n)$ $(j = 1, \ldots, n)$. For $j = 1$ and $j - 1$ we get the second largest eigenvalue $2\cos(2\pi/n)$ which is of multiplicity 2 and which corresponds to the second smallest eigenvalue $2 - 2\cos(2\pi/n)$ of the Laplacian. Two independent eigenvectors x, y are given by coordinates

$$x_l = \cos(2\pi l/n)\ (l = 1, \ldots, n), \quad y_l = \sin(2\pi l/n)\ (l = 1, \ldots, n).$$

If we represent vertices by points (x_i, y_i), the picture of the graph is a regular n-gon.

Graph representation obtained by the Laplacian matrix has been used in graph drawings [61,52,85].

Example. In our context interesting are also *fullerene graphs* corresponding to carbon compounds called *fullerenes*. Mathematically, fullerene graphs are planar regular graphs of degree 3 having as faces only pentagons and hexagons.[2] Although being planar, fullerene graphs are represented (and this really corresponds to actual positions of carbon atoms in a fullerene) in 3-space with its vertices embedded in a quasi-spherical surface.

A typical fullerene C_{60} is given in Fig. 1.

It can be described also as a truncated icosahedron and has the shape of a football.

Fullerene graphs have a nice 3D-representation in which the coordinates of the positions of vertices can be calculated from three eigenvectors of the adjacency matrix (the so called *topological coordinates* which were also used in producing the atlas [46]).

[2] It follows from the Euler theorem for planar graphs that the number of pentagons is exactly 12.

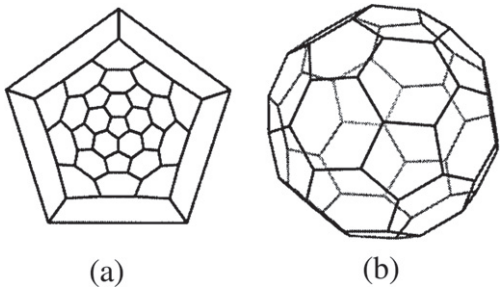

Fig. 1. (a) Planar and (b) 3D visualization of the icosahedral fullerene C_{60}.

Together with the Laplacian L and the normalized Laplacian $\hat{L}$ also the matrix $D^{-1}L$ has been used in clustering algorithms. According to [65] the last matrix performs best.

We shall mention a couple of other things.

Spectral filtering is an important method in handling huge sets of data. This method uses the eigenvectors of the adjacency and other graph matrices to find some clusters in data sets represented by graphs. For example, in [49] spectral filtering is applied in the study of Internet structure.

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 [76].

2.4. Computer vision and pattern recognition

Spectral graph theory has been widely applied to solve problems in the field of computer vision and pattern recognition. Examples include image segmentation, routing, image classification, etc. These methods use the spectrum, i.e., eigenvalues and eigenvectors, of the adjacency or Laplacian matrix of a graph.

The basic idea is to represent an image by a weighted graph with a vertex for each pixel and the edges between the neighboring pixels with weight depending on how similar the pixels are.

A more sophisticated idea is to represent an image's content by a graph with specially selected points as vertices. The interesting points are points in an image which have a well-defined position and can be robustly detected.

Several other graphs are used.

Graphs appearing in computer vision usually have a lot of vertices and simple eigenvalues [93].

Techniques from spectral-graph theory have been used to develop powerful algorithms in computer vision and pattern recognition. For instance, Shi and Malik [75] have shown how the Fiedler vector (i.e., the eigenvector associated to the second smallest eigenvalue of the Laplacian matrix) can be used to separate the foreground from the background structure in images. The original procedure from [45] 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.

Sarkar and Boyer [71] have shown how the eigenvector of the largest eigenvalue of a relevant[3] graph matrix can be used to group line segments.

[3] The global relationship among image features can be very effectively captured in the form of a graph whose vertices represent the image features and whose edges denote compatibility between the features. Two image features are said to be compatible if they exhibit pairwise organization, e.g., the two structures are of the same type, similar size, or have similar orientation. The edges in the graph are weighted according to the degree of compatibility between two vertices.

Graph clustering is an important issue in computer vision and pattern recognition, since graphs can be used for the high-level abstraction of scene and object structure.

Standard graph clustering methods need to solve the correspondence problems between vertices of the original and the transformed graph what could cause computational problems. Luo et al. [64,93] have proposed spectral invariants for graph clustering. These methods do not need to solve the vertex correspondence problem, instead they rely on using information from the spectrum of the Laplacian matrix.

An image, for example, of 100×100 pixels, each having 256 colour levels, can be represented as a point in $\mathbf{R}^{2560000}$. However, images in reality usually are contained in a subspace of much lower dimension. Several authors have explored the use of eigenvectors of the Laplacian and related operators to map data to a manifold in a low dimensional space [70,5]. These maps are similar as those described in graph clustering (Section 2.3).

Horaud and Sossa [54] have applied the spectral graph theory to image database indexing by comparing the coefficients of the polynomials of the Laplacian matrix of the weighted graph extracted from the image. This representation was used for indexing a large database of line drawings.

Luo et al. [64] have explored how ideas from spectral graph theory can be used to construct pattern spaces for sets of graphs. The idea has been to extract features that are permutation invariants from the adjacency matrices of the graphs under study, such as permutation invariant polynomials from the eigenvectors of the Laplacian matrix [93]. Pattern spaces may then be constructed from the feature vectors using techniques such as *principal components analysis* (PCA, procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called *principal components*).

2.5. Internet search

Web search engines are based on eigenvectors of the adjacency and some related graph matrices. The most known systems are PageRank [12] (used in Google) and Hyperlinked Induced Topics Search (HITS) [58].

The structure of the Internet is represented in this context by a digraph G where web pages correspond to vertices and links between the pages (hyperlinks) to arcs.

HITS exploits eigenvectors belonging to the largest eigenvalues of the matrices AA^T and A^TA where A is the adjacency matrix of a subgraph of G induced by the set of web pages obtained from search key words by some heuristics. The obtained eigenvectors define a certain ordering of selected web pages.

PageRank uses similar ideas. Random walks are considered in this model. In fact, the adjacency matrix of G is normalized so that the sum of entries in each row is equal to 1. This is achieved by dividing[4] the entries in each row by the out-degree of the corresponding vertex. Equivalently, we form a new matrix $P = D_+^{-1}A$ where D_+ is the diagonal matrix of out-degrees. The matrix P is a transition matrix of a Markov chain and the normalized eigenvector of the largest eigenvalue of its transpose P^T defines the steady-state of the chain. Pages are ranked by the coordinates of this eigenvector.

Expository paper [62] contains a survey of both techniques.

There are many papers in Computer Science literature on different aspects of using eigenvectors in Internet search engines.

From the mathematical point of view, the subject of ranking individuals or objects by eigenvectors of suitably chosen graph matrices is very old. One of the basic references is the thesis [92]. In particular, the ranking of the participants of a round-robin tournament can be carried out in that way (see, for example [28, p. 226]). These methods have been used in the sociology for a long time as well (see, for example [8]).

We reproduce here a relevant result. The following theorem of Wei [92] is noted in [31, p. 26]:

[4] Prior to this transformation, arcs going to all other vertices are added to each vertex without outgoing arcs. Also, in order to ensure primitivity of the matrix, at least one odd cycle is artificially formed if such one did not exist.

Theorem. *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 $1, 2, \ldots, n$. 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 $(s_k(1), s_k(2), \ldots, s_k(n))^T$ tends towards the eigenvector corresponding to the index of G.*

Counting walks with specified properties in a graph (or digraph) is related to graph spectra by the following well-known result (see [28, p. 44]).

Theorem. *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.*

Thus, for example, the number of closed walks of length k is equal to the kth spectral moment, since $\sum_{i=1}^{n} a_{ii}^{(k)} = \mathrm{tr}(A^k) = \sum_{i=1}^{n} \lambda_i^k$.

2.6. Load balancing and multiprocessor interconnection networks

The job which has to be executed by a multiprocessor system is divided into parts that are given to particular processors to handle them. We can assume that the whole job consists of a number of elementary jobs (items) so that each processor gets a number of such elementary jobs to execute. Mathematically, elementary jobs distribution among processors can be represented by a vector x whose coordinates are non-negative integers. Coordinates are associated to graph vertices and indicate how many elementary jobs are given to corresponding processors.

Vector x is usually changed during the work of the system because some elementary jobs are executed while new elementary jobs are permanently generated during the execution process. Of course, it would be optimal that the number of elementary jobs given to a processor is the same for all processors, i.e., that the vector x is an integer multiple of the vector j whose all coordinates are equal to 1. Since this is not always possible, it is reasonable that processors with a great number of elementary jobs send some of them to adjacent processors so that the job distribution becomes uniform if possible. In this way the so called problem of *load balancing* is important in managing multiprocessor systems. The load balancing problem requires creation of algorithms for moving elementary jobs among processors in order to achieve the uniform distribution.

We shall present an algorithm for the load balancing problem which is based on the Laplacian matrix of a graph.

Let G be a connected graph on n vertices. Eigenvalues and corresponding orthonormal eigenvectors of the Laplacian $L = D - A$ of G are denoted by $\nu_1, \nu_2, \ldots, \nu_n = 0$ and $u_1, u_2, \ldots, u_n$, respectively. Any vector x from $\mathbf{R}^n$ can be represented as a linear combination of the form $x = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n$.

Suppose now that G has distinct Laplacian eigenvalues $\mu_1, \mu_2, \ldots, \mu_m = 0$ with multiplicities $k_1, k_2, \ldots, k_m = 1$, respectively. Vector x can now be represented in the form $x = y_1 + y_2 + \cdots + y_m$ where y_i belong to the eigenspace of μ_i for $i = 1, 2, \ldots, m$. We also have $y_m = \beta j$ for some β.

Since $Lx = L(y_1 + y_2 + \cdots + y_m) = \mu_1 y_1 + \mu_2 y_2 + \cdots + \mu_m y_m$, we have $x^{(1)} = x - \frac{1}{\mu_1} Lx = \left(I - \frac{1}{\mu_1} L\right) x = \left(1 - \frac{\mu_2}{\mu_1}\right) y_2 + \cdots + \beta j$. We see that the component of x in the eigenspace of μ_1 has been canceled by the transformation by the matrix $I - \frac{1}{\mu_1} L$ while the component in the eigenspace of μ_m remains unchanged. The transformation $I - \frac{1}{\mu_2} L$ will cause that the component of $x^{(2)} = \left(I - \frac{1}{\mu_2} L\right) x^{(1)}$ in the eigenspace of μ_2 disappears. Continuing in this way

$$x^{(k)} = \left(I - \frac{1}{\mu_k} L\right) x^{(k-1)}, \quad k = 1, 2, \ldots, m-1 \tag{1}$$

we shall obtain $x^{(m-1)} = \beta j$.

We have seen how a vector x can be transformed to a multiple of j using the iteration process (1) which involves the Laplacian matrix of the multiprocessor graph G. It remains to see what relations (1) mean in terms of load moving.

Let vector $x^{(k)}$ have coordinates $x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}$. Relations (1) can be rewritten in the form

$$x_i^{(k)} = x_i^{(k-1)} - \frac{1}{\mu_k} \sum_{i*j} \left(d_i x_i^{(k-1)} - x_j^{(k-1)} \right) \tag{2}$$

where d_i is the degree of vertex i. This means that the current load at vertex i is changed in such a way that vertex (processor) i sends $\frac{1}{\mu_k}$th part of its load to each of its d_i neighbors and, because this holds for every vertex, also receives $\frac{1}{\mu_k}$th part of the load from each of its d_i neighbors. The amounts should be added algebraically and in this way we get final value of the flow through edge ij (sent either from i to j or vice versa). The obtained flow is ℓ_2-optimal, i.e., the sum of the squares of particular edge flows is minimal (see, for example [42]).

The number of iterations in (1) is equal to the number of non-zero distinct Laplacian eigenvalues of the underlying graph. In addition, maximum vertex degree Δ of G also affects computation of the balancing flow. Therefore, the complexity of the balancing flow calculations essentially depends on the product $m\Delta$ and that is why this quantity was proposed in [42] as a parameter relevant for the choice and the design of multiprocessor interconnection networks.

See references [39,40,48,55,56] for further information on the load balancing problem.

As we have already pointed out, the graph invariant obtained as the product of the number of distinct eigenvalues m and the maximum vertex degree Δ of G has been investigated in [42] related to the design of multiprocessor topologies. The main conclusion of [42] with respect to the multiprocessor design and, in particular to the load balancing within given multiprocessor systems was the following: if $m\Delta$ is small for a given graph G, the corresponding multiprocessor topology was expected to have good communication properties and has been called *well-suited*. The graphs with large $m\Delta$ were called *ill-suited* and were not considered suitable for design of multiprocessor interconnection networks.

The following definitions of four kinds of graph *tightness* have been introduced and used in [23–25].

First type mixed tightness $t_1(G)$ of a graph G is defined as the product of the number of distinct eigenvalues m and the maximum vertex degree Δ of G, i.e., $t_1(G) = m\Delta$.

Structural tightness $stt(G)$ is the product $(D + 1)\Delta$ where D is diameter and Δ is the maximum vertex degree of a graph G.

Spectral tightness $spt(G)$ is the product of the number of distinct eigenvalues m and the largest eigenvalue λ_1 of a graph G.

Second type mixed tightness $t_2(G)$ is defined as a function of the diameter D of G and the largest eigenvalue λ_1, i.e., $t_2(G) = (D + 1)\lambda_1$.

Several arguments were given which support the claim that graphs with small tightness t_2 are well suited for multiprocessor interconnection networks.

It was proved that the number of connected graphs with a bounded tightness is finite and graphs with tightness values not exceeding 9 are determined explicitly. There are 69 such graphs and they contain up to 10 vertices. In addition, graphs with minimal tightness values when the number of vertices is $n = 2, \ldots, 10$ are identified.

2.7. Anti-virus protection vs. spread of knowledge

The largest eigenvalue λ_1 of the adjacency matrix plays an important role in modeling virus propagation in computer networks. The smaller the largest eigenvalue, the larger the robustness of a network against the spread of viruses. In fact, it was shown in [91] that the epidemic threshold in spreading viruses is proportional to $1/\lambda_1$. Another model of virus propagation in computer networks has been developed in [88] with the same conclusion concerning $1/\lambda_1$.

Motivated by the above facts, the authors of [38] determine graphs with minimal λ_1 among graphs with given number of vertices and having a given diameter.

Research and development networks (R&D networks) are studied using the largest eigenvalue of the adjacency matrix in [59,60]. In such networks it is desirable that the knowledge is spread through network as much as possible. Therefore the tendency is to achieve high values of the largest eigenvalue, just opposite to considerations of virus propagation.

An intuitive explanation of both phenomena, advantage to have minimal value of λ_1 for virus protection and maximal value of λ_1 for knowledge spread, can be obtained by the fact that the number of walks of length k in a connected graph behaves asymptotically as $c\lambda_1^k$ for a constant $c > 0$. The greater the number of walks the more intensive is the spread of the mowing substance, does not matter whether this is the virus or the knowledge.

The virus propagation model established in [91] is a discrete time model. It uses the vector $\mathbf{P}_t = (p_{1,t}, p_{2,t}, \ldots, p_{n,t})^T$ where $p_{i,t}$ is the probability that the vertex i is infected at time t. The basic relation is

$$p_{i,t} = (1-\delta)p_{i,t-1} + \beta \sum_{i\sim j} p_{j,t-1},$$

where β is the virus birth rate on an edge connected to an infected vertex and δ the virus curing rate on an infected vertex. The corresponding matrix relation is

$$\mathbf{P}_t = ((1-\delta)I + \beta A)\mathbf{P}_{t-1},$$

where A is the adjacency matrix of the graph representing the network.

Further we have $\mathbf{P}_t = S^t\mathbf{P}_0$ for $t = 0, 1, 2, \ldots$ where $S = (1-\delta)I + \beta A$. We see that the the vector $\mathbf{P}_t$ will tend to a zero-vector for $t \to \infty$ if and only if all eigenvalues of the matrix S are smaller than 1 in modulus. This would mean that virus epidemic has died and will happen if $1 - \delta + \beta\lambda_1 < 1$, i.e.,

$$\frac{\beta}{\delta} < \frac{1}{\lambda_1}.$$

Given β and δ, we see that the network is as safer as the smaller is λ_1. We can denote the quantity $\tau = \frac{1}{\lambda_1}$ as the epidemic threshold in spreading viruses. Hence if $\frac{\beta}{\delta} < \tau$ the network is safe and in the opposite case the network will be conquered by viruses.

We shall describe the model of R&D networks according to [59,60]. Consider an industry in which firms engage in pairwise R&D collaborations with other firms. Collaborations allow the growth of knowledge within the firm and an increase in the probability to introduce innovations that yield profit to the firm. R&D network is represented by a graph G in which the firms are vertices and collaborations between firms are edges. Let x_i be the level of knowledge of the firm i and it depends on time t. It is assumed that the growth of knowledge within a firm is proportional to the sum of levels of knowledge of its neighbors. This leads to a system of linear differential equations and to the conclusion that the expected profit per unit of time is equal to the largest eigenvalue of the relevant component of G minus the costs of formation and maintaining the collaboration links with other firms.

There are numerous mathematical investigations in both directions: to find graphs in particular classes of graphs which have minimal or maximal largest eigenvalue. We mention a few results and references.

We need the following definition.

Definition. A graph G with the edge set E_G is called a *nested split graph* if its vertices can be ordered so that $jq \in E_G$ implies $ip \in E_G$ whenever $i \leqslant j$ and $p \leqslant q$.

This definition is used in [31], where the graphs in question were called graphs with a *stepwise* adjacency matrix. Some other definitions and terms are used in the literature, e.g., degree maximal graphs, threshold graphs. Note that graphs with a stepwise adjacency matrix are exactly the nested split graphs. We also have

Proposition. *A graph is a nested split graph if and only if it does not contain as an induced subgraph any of the graphs* P_4, $2K_2$, C_4.

It is well-known that in connected graphs with the given numbers of vertices and edges the graph with maximal largest eigenvalue is a nested split graph (see, for example [31, pp. 72–73]).

There are less results concerning minimal values of the largest eigenvalue. See Section 3 for a survey and some recent results.

2.8. Statistical databases and social networks

Statistical databases are those that allow only statistical access to their records. Individual values are typically deemed confidential and are not to be disclosed, either directly or indirectly. Thus, users of a statistical database are restricted to statistical types of queries, such as looking for the sum of values, minimum or maximum value of some records, etc. Moreover, no sequence of answered queries should enable a user to obtain any of the confidential individual values. However, if a user is able to reveal a confidential individual value, the database is said to be *compromised*. Statistical databases that cannot be compromised are called *compromise-free* or *secure*.

We shall concentrate on the restricted case where queries are related to the sum of values of records in the database and each record is contained in at most 2 queries. Then the query matrix corresponds to a an incidence matrix of a graph G, where queries correspond to vertices and records correspond to edges.

Surprisingly, the results from [9,11] show an interesting connection between compromise-free query collections and graphs with least eigenvalue -2 [33]. This connection was recognized in the paper [10].

A *dumbbell* is a graph obtained by joining two cycles by a path (possibly of length 0). If both cycles of a dumbbell are odd, then the dumbbell is called *odd*. It is nowadays well-known (cf., e.g. [33, p. 126]) that eigenvectors of a basis of the eigenspace of the eigenvalue -2 in line graphs can be obtained by certain edge valuations of even cycles and odd dumbbells in the root graphs.

The following theorem was proved in [9,11] with another terminology. See next section for some bibliographical comments.

Theorem. *A database is compromise-free if and only if each edge of G is contained either in an even cycle or in an odd dumbbell of G.*

The problem of protecting the privacy appears also in *social networks* at 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 false edges in such a way that global characteristics of the network are unchanged. This is achieved 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 [94].

2.9. Quantum computing

Quantum computation is a model of computation based on the principles of quantum mechanics although the corresponding computers have not yet been realized. In spite of the non-existence of actual machines, the theory of quantum computing is very much developed. For a general overview on Quantum Information Technology see, for example, special issue of the journal NEC Research & Developments [69].

A graph is called *integral* if its spectrum consists entirely of integers. Integral graphs have been studied for decades as a kind of mathematical curiosity without any idea for what they could be used outside mathematics (see [1] for a survey of results up to 2004). In particular, there are exactly 13 connected, cubic, integral graphs [13]. Among them are, for example, the 3-dimensional cube and the Petersen graph.

It has been discovered recently [17] that integral graphs can play a role in the so called perfect state transfer in quantum spin networks. Speaking in terms of quantum physics, there is perfect state

transfer between two vertices of a graph if a single excitation can travel with fidelity one between the corresponding sites of a spin system modeled by the graph [74].

Let G be a graph with adjacency matrix A and consider the matrix $H(t) = e^{iAt}$ where t is a real variable and $i^2 = -1$. According to [50], perfect state transfer occurs between vertices u and v of G if there is a value of t such that $|H(t)_{u,v}| = 1$. This can happen in integral graphs but not always.

Further details on this topic can be found in [73,50,51,74,84].

The 3-dimensional cube is the only connected cubic integral graph with perfect state transfer [74]. Some other results in this direction have been obtained in [2,3].

3. Comments and new results

After completing our survey of applications of graph spectra to several branches in Computer Science, we have some observations and comments including some new results.

Our general suggestion is that mathematicians should react on the explosion of the number of papers in Computer Science which use graph spectra by selecting for their own research some subjects from or inspired by such applications.

We have a number of detailed comments.

Ad. 1. We shall announce some results on graphs with a minimal value of the largest eigenvalue. According to Section 2.7, these graphs are models of computer networks resistent to the spread of viruses.

Graphs with a minimal value of the largest eigenvalue in a set of graphs will be called *minimal* graphs.

Let $\deg(v)$ be the degree of the vertex v. An *internal path* in some graph is a path $v_0, v_1, \ldots, v_{k+1}$ for which $\deg(v_0), \deg(v_{k+1}) \geqslant 3$ and $\deg(v_1) = \cdots = \deg(v_k) = 2$ (here $k \geqslant 0$, or $k \geqslant 2$ whenever $v_{k+1} = v_0$).

Consider connected graphs with fixed numbers of vertices and edges.

A minimal graph cannot contain vertices of degree 1 because deleting a vertex of degree 1 and a simultaneous insertion of a vertex of degree 2 in the middle of an edge on an internal path would diminish the largest eigenvalue λ_1 (by a result from [53]). As a consequence, all edges belong to internal paths. Subdividing edges in such graphs we can further diminish λ_1. However, we have $\lambda_1 \geqslant \Delta/\sqrt{\Delta - 1}$, where Δ is the maximum vertex degree (as follows form a result from [53]). Therefore it is reasonable to choose Δ as small as possible. Obviously we should accept $\Delta = 3$ since graphs with $\Delta < 3$ are not of much interest in this context.

Let H be a connected graph with q selected edges, labeled by $1, 2, \ldots, q$, and with vertex degrees at least 3. For each $i = 1, 2, \ldots, q$ we want to subdivide edge i by inserting l_i vertices of degree 2. Suppose that $\Sigma_{i=1}^{q} l_i = e$ is fixed. In this way we obtain a graph G. An explicit relation between the largest eigenvalue of G and quantities $l_1, l_2, \ldots, l_q$ has been found in [79]. A standard procedure with Lagrange's multipliers for finding the minimum of an implicit function leads to the conclusion that under some reasonable additional assumptions the quantities l_i should be almost equal (equal if possible). Such a subdivision of a graph is called *balanced* subdivision.

Indeed many of the examples of minimal graphs found in [38] by computer have this property.

Earlier results on the subject are special cases of this result. It started with a conjecture [20] that a balanced subdivision minimizes the largest eigenvalue in a cycle with an additional edge. The problem was solved in a generalized form in [80]. The paper [78] solves the problem for bicyclic graphs. See also [6] for broken wheels and [4] for trees.

The above presented procedure enables explicit construction of minimal tricyclic graphs [79]. Previous result for bicyclic graphs has been proved in a shorter way.

Our conclusion is that balanced subdivisions of cubic graphs should be considered as good models of virus resistent computer networks.

Ad. 2. In problems of Section 2.7 we can use the signless Laplacian matrix instead of the adjacency matrix. Then instead of the largest eigenvalue λ_1 of the adjacency matrix the largest eigenvalue q_1 of the signless Laplacian becomes relevant. To explain the situation the standard definition of a walk (used in Section 2.5) will be slightly modified.

Definition. *A walk (of length k) in an (undirected) graph G is an alternating sequence* $v_1, e_1, v_2, e_2, \ldots, v_k, e_k, v_{k+1}$ *of vertices* $v_1, v_2, \ldots, v_k, v_{k+1}$ *and edges* $e_1, e_2, \ldots, e_k$ *such that for any* $i = 1, 2, \ldots, k$ *the vertices* v_i *and* v_{i+1} *are distinct end-vertices of the edge* e_i.

Such a walk can be imagined as an actual walk of a traveler along the edges in a diagrammatic representation of the graph under consideration. The traveler always walks along an edge from one end-vertex to the other. Suppose now that we allow the traveler to change his mind when coming to the midpoint of an edge: instead of continuing along the edge towards the other end-vertex, he could return to the initial end-vertex and continue as he wishes. Then the basic constituent of a walk is no longer an edge; rather we could speak of a walk as a sequence of *semi-edges*. Such walks could be called *semi-edge walks*. A semi-edge in a walk could be followed by the other semi-edge of the same edge (thus completing the edge) or by the same semi-edge in which case the traveler returns to the vertex at which he started. A formal definition of a semi-edge walk is obtained from the above definition of a walk by deleting the word "distinct" from the description of end-vertices. Hence we have the following definition.

Definition. *A semi-edge walk (of length k) in an (undirected) graph G is an alternating sequence* $v_1, e_1, v_2, e_2, \ldots, v_k, e_k, v_{k+1}$ *of vertices* $v_1, v_2, \ldots, v_{k+1}$ *and edges* $e_1, e_2, \ldots, e_k$ *such that for any* $i = 1, 2, \ldots, k$ *the vertices* v_i *and* v_{i+1} *are end-vertices (not necessarily distinct) of the edge* e_i.

In both definitions we shall say that the walk *starts* at the vertex v_1 and *terminates* at the vertex v_{k+1}.

The well known theorem concerning the powers of the adjacency matrix given in Section 2.5, has the following counterpart for the signless Laplacian (see [35]).

Theorem. *Let Q be the signless Laplacian of a graph G. The* (i, j)*-entry of the matrix* Q^k *is equal to the number of semi-edge walks of length k starting at vertex i and terminating at vertex j.*

The number of semi-edge walks of length k in a connected graph behaves asymptotically as cq_1^k for a constant $c > 0$, where q_1 is the largest eigenvalue of the signless Laplacian.

Both standard and semi-edge walks can be considered as *random walks*. This situation appears if we prescribe that a walk at each point (vertex for standard walks and vertex or the midpoint of an edge for semi-edge walks) can be continued with equal probabilities. This means that a semi-edge walk will remain at the same vertex (after having passed two semi-edges) with probability $1/2$. Such random walks are known in the literature as *lazy random walks* (see, for example [82]).

Based on these observations, we could suggest to try to create models for virus propagation and the spread of knowledge in which the adjacency matrix would be replaced by the signless Laplacian. In such models desirable graphs for anti-virus protection would be those with small q_1 and for R&D networks those with large q_1. We believe that there are situations in which viruses or knowledge move along lazy random walks rather than along standard random walks. This can be expected in situations when the vertices when receiving something from their neighbors are likely to respond back with some action.

Like in the case of the adjacency matrix, in connected graphs with the given numbers of vertices and edges the graph with maximal largest signless Laplacian eigenvalue is again a nested split graph [35], not necessarily the same as for the adjacency matrix. There are almost no results on graphs with minimal largest signless Laplacian eigenvalue. One exception appears in [37] where the result for bicyclic graphs [78] for the adjacency matrix has been repeated for the signless Laplacian.

Ad. 3. For non-trivial connected graphs the matrices $D^{-1}A$ and $(2D)^{-1}Q = (2D)^{-1}(D + A) = \frac{1}{2}(I + D^{-1}A)$ are transition matrices of Markov chains for random and lazy random walks.

Note that the normalized Laplacian matrix $\hat{L} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}} = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ and the normalized signess Laplacian matrix $\hat{Q} = D^{-\frac{1}{2}}QD^{-\frac{1}{2}} = D^{-\frac{1}{2}}(D + A)D^{-\frac{1}{2}} = I + D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$

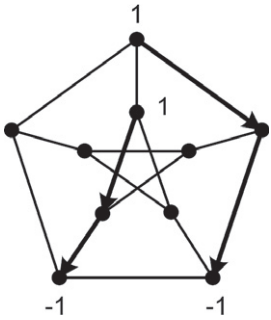

Fig. 2. An eigenvector for eigenvalue 1 and a load balancing flow.

are connected by the relation $\hat{Q} = -\hat{L} + 2I$. This means that $\hat{Q}$-theory is simply reduced to $\hat{L}$-theory. A similar statement holds for the matrix $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$.

The book [28] describes some spectral properties of the matrices $D^{-1}A$ and $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ (they are similar, see p. 48) and $\frac{1}{2}(I + D^{-1}A)$ (see p. 110). All three matrices have real eigenvalues.

Ad. 4. As defined in Section 2.9, a graph is called *integral* if its spectrum consists entirely of integers. Each eigenvalue has integral eigenvectors and each eigenspace has a basis consisting of such eigenvectors.

In integral graphs load balancing algorithms, which use eigenvalues and eigenvectors, can be executed in integer arithmetics as noted in [24]. In addition, these integral eigenvectors can be selected so that they contain a lot of coordinates equal to 0 [26].

As explained in Section 2.6, any load distribution vector can be represented as a linear combination of eigenvectors.

Recall that $3, 1^5, (-2)^4$ is the spectrum of the Petersen graph. An eigenvector for eigenvalue 1 is given in Fig. 2.

We can introduce several load balancing schemes on the set of selected integral eigenvectors. In Fig. 2 we see a load balancing flow which occupies a minimal number of arcs. At this step we can introduce various additional optimality criteria.

For more details see [26].

The further study of integral graphs in connection to multiprocessor topologies seems to be a promising subject for future research.

4. Bibliographical notes

We shall first present some expository texts on applications of graph spectra.

The books [28,27] contain each a chapter on applications of graph eigenvalues.

The book [34] also contains a chapter on applications. There are sections on Physics, Chemistry, Computer Sciences and Mathematics itself.

According to its Preface, the purpose of the book [30] is to draw the attention of mathematical community to rapidly growing applications of the theory of graph spectra. Besides classical and well documented applications to Chemistry and Physics, we are witnesses of the appearance of graph eigenvalues in Computer Science in various investigations. There are also applications in several other fields like Biology, Geography, Economics and Social Sciences. A monograph with a comprehensive treatment of applications of graphs spectra is missing at the present.

The book [30] contains five chapters: an introductory chapter with a survey of applications by representative examples and four case studies (one in Computer Science and three in Chemistry).

The introductory text [21] provides an introduction to the theory of graph spectra and a short survey of applications of graph spectra together with some selected bibliographies on applications. We have mentioned applications to Chemistry, Physics, Computer Sciences and Mathematics itself.

Graph spectra are used in many other branches of science including Biology, Geography, Economics and Social Sciences.

Referring to the book [28] as "the current standard work on algebraic graph theory", Van Mieghem gave in his book [86] a 20 page appendix on graph spectra, thus pointing out the importance of this subject for communications networks and systems.

The paper [82] is a tutorial on the basic facts of the theory of graph spectra and its applications in Computer Science delivered at the 48th Annual IEEE Symposium on Foundations of Computer Science.

Next, we shall give some details on the independent discoveries of the the structure of the eigenspace of -2 in line graphs related to statistical data bases. It is interesting that original Doob's description [41] in 1973 of the eigenspace of -2 in line graphs in terms of even cycles and odd dumbbells has been extended to generalized line graphs by Cvetković et al. [29] in 1981 in terms of the chain groups, not explicitly dealing with cycles and dumbbells. The independent discovery of Branković, Miller and Širáň [11] in 1996 put implicitly some light on the description of the eigenspace in generalized line graphs a bit before Cvetković et al. in 2001 (the paper [32] was submitted in 1998), using the star complement technique and without being aware of [11], gave the entire description of the eigenspace.

Acknowledgement

We are grateful to Lj. Trajković, T. Aleksić and D. Stevanović for some references and information related to Sections 2.2, 2.4 and 2.9, respectively.

References

[1] K. Balińska, D. Cvetković, Z. Radosavljević, S. Simić, D. Stevanović, A survey on integral graphs, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. 13 (2002) 42–65, Errata, These Publications, 15 (2004) 112.
[2] M. Bašić, M. Petković, Some classes of integral circulant graphs allowing and not allowing perfect state transfer, Appl. Math. Lett. 22 (2009) 1609–1615.
[3] M. Bašić, M. Petković, D. Stevanović, Perfect state transfer in integral circulant graphs, Appl. Math. Lett. 22 (2009) 1117–1121.
[4] F. Belardo, E.M. Li Marzi, S.K. Simić, Trees with minimal index and diameter at most four, Discrete Math. 310 (2010) 1708–1714.
[5] M. Belkin, p. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (2000) 1373–1396.
[6] F.K. Bell, S.K. Simić, On the index of broken wheels, Linear and Multilinear Algebra 39 (1995) 137–152.
[7] N.L. Biggs, Algebraic Graph Theory, Cambridge University Press, Cambridge, 1993.
[8] P. Bonacich, Power and centrality: A family of measures, Amer. J. Soc. 92 (1987) 1170–1182.
[9] L. Branković, Usability of secure statistical data bases, Ph.D. Thesis, Newcastle, Australia, 1998.
[10] Lj. Branković, D. Cvetković, The eigenspace of the eigenvalue -2 in generalized line graphs and a problem in security of statistical data bases, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. 14 (2003) 37–48.
[11] L. Branković, M. Miller, J. Širáň, Graphs, (0,1)-matrices and usability of statistical data bases, Congr. Numer. 120 (1996) 186–192.
[12] S. Brin, L. Page, The Anatomy of Large-Scale Hypertextual Web Search Engine, in: Proceedings of the Seventh International WWW Conference, 1998.
[13] F.C. Bussemaker, D. Cvetković, There are exactly 13 connected, cubic, integral graphs, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. 544-576 (1976) 43–48.
[14] H. Chang, R. Govindan, S. Jamin, S. Shenker, W. Willinger, Towards capturing representative AS-level Internet topologies, in: Proceedings of the ACM SIGMETRICS Performance Evaluation Review, vol. 30(1), New York, NY, June 2002, pp. 280–281.
[15] J. Chen, L. Trajković, Analysis of Internet topology data, in: Proceedings of the IEEE International Symponisum, Circuits and Systems, ISCAS 2004, Vancouver, BC, May 2004, pp. 629–632.
[16] Q. Chen , H. Chang, R. Govindan, S. Jamin, S. Shenker, W. Willinger , The origin of power laws in Internet topologies revisited, in: Proceedings of INFOCOM, New York, NY, USA, April 2002, pp. 608–617.
[17] M. Christandl, N. Datta, A. Ekert, A.J. Landahl, Perfect state transfer in quantum spin networks, Phys. Rev. Lett. 92 (2004) 187902.
[18] F. Chung, Spectral Graph Theory, American Mathematical Society, Providence, Rhode Island, 1997.
[19] F. Chung, L. Lu, Complex Graphs and Networks, American Mathematical Society, Providence, Rhode Island, 2006.
[20] D. Cvetković, Problems 2, 3 (Graphs, Hypergraphs and Applications, in: H. Sachs (Ed.), Proceedings of the Conference Graph Theory Held in Eyba, October 1984, Teubner, Leipzig, 1985, p. 211.
[21] D. Cvetković, Applications of Graph Spectra: An introduction to the literature, Applications of Graph Spectra, Zbornik radova 13(21), ed. D.Cvetković, I. Gutman, Mathematical Institute SANU, Belgrade, 2009, pp. 7–31.
[22] D. Cvetković, M. Čangalović and V. Kovačević-Vujčić, Semidefinite programming methods for the symmetric traveling salesman problem, in: *Integer Programming and Combinatorial Optimization, Proc. 7th Internat. IPCO Conf.*, Graz, Austria, June 1999 (G. Cornuejols, R. E. Burkard and G. J. Woeginger, eds.), Lecture Notes Comp.Sci.1610, Springer, Berlin (1999), pp. 126–136.
[23] D. Cvetković, D. Davidović, Applications of some graph invariants to the analysis of multiprocessor topologies, YUJOR 18 (2) (2008) 173–186.

[24] D. Cvetković, T. Davidović, Multiprocessor interconnection networks with small tightness, Internat. J. Found. Comput. Sci. 20 (5) (2009) 941–963.
[25] Cvetković D., Davidović T., *Multiprocessor interconnection networks*, *Applications of Graph Spectra*, Zbornik radova 13(21), ed. D.Cvetković, I.Gutman, Mathematical Institute SANU, Belgrade, 2009, 33–63.
[26] D. Cvetković, T. Davidović, A. Ilić, S.K. Simić, Graphs for small multiprocessor interconnection networks, Appl. Math. Comput. 217 (2010) 2468–2480.
[27] D. Cvetković, M. Doob, I. Gutman, A. Torgašev, Recent Results in the Theory of Graph Spectra, North-Holland, Amsterdam, 1988.
[28] D. Cvetković, M. Doob, H. Sachs, Spectra of Graphs, Theory and Application, third ed., Johann Ambrosius Barth Verlag, Heidelberg–Leipzig, 1995.
[29] D. Cvetković, M. Doob, S. Simić, Generalized Line Graphs, J. Graph Theory 5 (4) (1981) 385–399.
[30] D. Cvetković, I. Gutman, (Eds.,), *Applications of Graph Spectra*, Zbornik radova 13(21), Mathematical Institute SANU, Belgrade, 2009.
[31] D. Cvetković, P. Rowlinson, S.K. Simić, Eigenspaces of Graphs, Cambridge University Press, Cambridge, 1997.
[32] D. Cvetković, P. Rowlinson, S.K. Simić, Graphs with least eigenvalue -2: the star complement technique, J. Algebraic Combin. 14 (2001) 5–16.
[33] D. Cvetković, P. Rowlinson, S.K. Simić, Spectral Generalizations of Line Graphs, on Graphs with Least Eigenvalue -2, Cambridge University Press, Cambridge, 2004.
[34] D. Cvetković, P. Rowlinson, S.K. Simić, An Introduction to the Theory of Graph Spectra, Cambridge University Press, Cambridge, 2009.
[35] D. Cvetković, P. Rowlinson, S.K. Simić, Signless Laplacians of finite graphs, Linear Algebra Appl. 423 (1) (2007) 155–171.
[36] Cvetković D., Simić S., *The second largest eigenvalue of a graph – a survey*, Filomat, 9 (1995), No.3, Int. Conf. Algebra, Logic Discrete Math., Niš, April 14–16, 1995, (ed. S. Bogdanović, M. Ćirić, Ž. Perović), 449–472.
[37] D. Cvetković, S.K. Simić, Towards a spectral theory of graphs based on the signless Laplacian, II, Linear Algebra Appl. 432 (2010) 2257–2272.
[38] E.R. van Dam, R.E. Kooij, The minimal spectral radius of graphs with a given diameter, Linear Algebra Appl. 423 (2007) 408–419.
[39] Decker T., Monien B., Preis R., *Towards optimal load balancing topologies*, *Proc. 6th EuroPar Conf.*, Lect. Notes Comput. Sci. 1900, Springer, Berlin, 277–287.
[40] R. Dickmann, A. Frommer, B. Monien, Efficient schemes for nearest neighbor load balancing, Parallel Comput. 25 (1999) 789–812.
[41] M. Doob, An interrelation between line graphs, eigenvalues, and matroids, J. Combin. Theory Ser. B 15 (1973) 40–50.
[42] R. Elsässer, R. Královič, B. Monien, Sparse topologies with small spectrum size, Theor. Comput. Sci. 307 (2003) 549–565.
[43] Faloutsos M., Faloutsos P., Faloutsos C., *On power-low relationships of the Internet topology*, Proc. ACM SIGCOMM '99, ACM Press, New York, 1999, 251–262.
[44] M. Fiedler, Algebraic connectivity of graphs, Czechoslovak J. Math. 23 (98) (1973) 298–305.
[45] M. Fiedler, A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory, Czechoslovak J. Math. 25 (100) (1975) 619–633.
[46] P.W. Fowler, D.E. Manolopoulos, An Atlas of Fullerenes, Clarendon Press, Oxford, 1995.
[47] Gago S., *Spectral techniques in complex networks*, *Selected Topics on Applications of Graph Spectra*, Zbornik radova, ed. D.Cvetković, I.Gutman, Mathematical Institute SANU, Belgrade, to appear.
[48] Ghosh B., Muthukrishnan S., Schultz M. H., *First and second order diffusive methods for rapid, coarse, distributed load balancing*, *Proc. 8th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'96)*, ACM Press, 1996, 72–81; Theory Comput. Systems, 31(1998), 331–354.
[49] Gkantsisis C., Mihail M., Zegura E., *Spectral analysis of internet topologies*, Proc. INFOCOM2003, San Francisco, CA, Vol. 1, march 2003, 364–374.
[50] Godsil C., *Periodic graphs*, arXiv: 0806.2704 [math.CO].
[51] Godsil C., *When can perfect state transfer occur ?*, arXiv: 1001.0231v1 [math.CO].
[52] K.M. Hall, An r-dimensional quadratic placement algorithm, Management Sci. 17 (1970) 219–229.
[53] A.J. Hoffman, J.H. Smith, On the spectral radii of topologically equivalent graphs, in: M. Fiedler (Ed.), Recent Advances in Graph Theory, Academia Praha, 1975, pp. 273–281.
[54] R. Horaud, H. Sossa, Polyhedral object recognition by indexing, Pattern Recognition 28 (12) (1995) 1855–1870.
[55] Y.F. Hu, R.J. Blake, An improved diffusion algorithm for dymanic load balancing, Parallel Comput. 25 (1999) 417–444.
[56] Y.F. Hu, R.J. Blake, D.R. Emerson, An optimal migration algorithm for dymanic load balancing, Concurrency Pract. Exp. 10 (1998) 467–483.
[57] Kar S., Moura J.M.F., *Ramanujan topologies for decision making in sensor networks*, 44th Annual Allerton Conference on Communications, Control, and Computing, Allerton, Monticello, IL, Sep. 27-29, 2006.
[58] J. Kleinberg, Authoratitive sources in a hyperlinked environment, J. ACM 48 (1999) 604–632.
[59] König, M.D., Battiston S., Napoletano M., Schweityer F., *The efficiency and evolution of R&D networks*, Working Paper 08/95, Economics Working Paper Series, Eidgenössische Technische Hochschule Zürich, Zïich, 2008.
[60] M.D. König, S. Battiston, M. Napoletano, F. Schweityer, On algebraic graph theory and the dynamics of innovation networks, Networks and Heterogenous Media 3 (2) (2008) 201–219.
[61] Y. Koren, Drawing graphs by eigenvectors: theory and practice, Comput. Math. Appl. 49 (2005) 1867–1888.
[62] A.N. Langville, C.D. Meyer, A survey of eigenvector methods for Web information retrieval, SIAM Rev. 47 (1) (2005) 135–161.
[63] A. Lubotzky, R. Phillips, P. Sarnak, Ramanujan graphs, Combinatorica 8 (1988) 261–277.
[64] B. Luo, R.C. Wilson, E.R. Hancock, Spectral embedding of graphs, Pattern Recognition 36 (2003) 2213–2230.
[65] U. von Luxburg, A tutorial on spectral clustering, Statist. Comput. 17 (2007) 395–416.
[66] Mihail M. Papadimitrou C.H., *On the eigenvalue power-low*, RANDOM 2002, LNCS 2483, Springer, Berlin, 2002, 254–262.
[67] B. Mohar, S. Poljak, Eigenvalues in combinatorial optimization, in: R. Brualdi, S. Friedland, V. Klee (Eds.), Combinatorial and Graph-Theoretical Problems in Linear Algebra, Springer-Verlag, New York, 1993, pp. 107–151.
[68] Najiminaini M., Subedi L., Trajković Lj., *Analysis of Internet topologies: a historical view*, *Proc. IEEE Int. Symp. Circuits and Systems*, Taipei, Taiwan, May 2009, 1697–1700.

[69] *Quantum Information Technology*, NEC Research & Developments, 44(3) (2003) 263–306.
[70] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[71] S. Sarkar, K.L. Boyer, Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors, Comput. Vision Image Understanding 71 (1998) 110–136.
[72] Sawilla R.,*A survey of data mining of graphs using spectral graph theory,* Defence R&D Canada ' Ottawa, Technical Memorandum TM 2008-317, Ottawa, 2008.
[73] N. Saxena, S. Severini, I.E. Shparlinski, Parameters of integral circulant graphs and periodic quantum dynamics, Internat. J. Quantum Inf. 5 (3) (2007) 417–430.
[74] Severini, S., The 3-dimensional cube is the only connected cubic graph with perfect state transfer, arXiv: 1001.0674v2 [quant-ph].
[75] Shi J., 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.
[76] A. Shokoufandeh, S.J. Dickinson, K. Siddiqi, S.W. Zucker, Indexing using a spectral encoding of topological structure, IEEE Trans. Comput. Vision Pattern Recognition 2 (1999) 491–497.
[77] G. Siganos, M. Faloutsos, P. Faloutsos, C. Faloutsos, Power-laws and the AS-level Internet topology, IEEE/ACM Trans. Netw. 11 (4) (2003) 514–524.
[78] S. Simić, On the largest eigenvalue of bicyclic graphs, Publ. Inst. Math. (Beograd) 46 (60) (1989) 1–6.
[79] S.K. Simić, Connected graphs of fixed order and size with minimal index, in preparation.
[80] S.K. Simić, V.Lj. Kocić, On the largest eigenvalue of some homeomorphic graphs, Publ. Inst. Math. (Beograd) 40 (54) (1986) 3–9.
[81] D.B. Skilicorn, Understanding Complex Data Bases: Data Mining with Matrix Decompositions, Chapman & Hall/CRC, New York, 2007.
[82] D.A. Spielman, Spectral Graph Theory and its Applications, in: 48th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2007, pp. 29–38.
[83] D.A. Spielman, S. Teng, Spectral partitioning works: planar graphs and finite element meshes, in: 37th Annual Symposium on Foundations of Computer Science, Burlington, VT, 1996, pp. 96–105.
[84] D. Stevanović, Applications of graph spectra in quantum physics, in: D. Cvetković, I. Gutman (Eds.), Selected Topics on Applications of Graph Spectra, Zbornik radova, Mathematical Institute SANU, Belgrade, in press.
[85] W.T. Tutte, How to draw a graph, Proc. London Math. Soc. 13 (1963) 743–768.
[86] P. Van Mieghem, Performance Analysis of Communications Networks and Systems, Cambridge University Press, Cambridge, 2006.
[87] P. Van Mieghem, Graph Spectra for Complex Networks, Cambridge University Press, Cambridge, 2010.
[88] P. Van Mieghem, J. Omić, R. Kooij, Virus spread in networks, IEEE/ACM Trans. Netw. 17 (2009) 1–14.
[89] D. Vukadinović, P. Huang, T. Erlebach, A spectral analysis of the Internet topology, Technical Report 118, TIK Institute, ETH Zürich, 2001.
[90] D. Vukadinović, P. Huang, T. Erlebach, On the spectrum and structure of Internet topology graphs, in: Proceedings of the Second International Workshop on Innovative Internet Computing systems, IICS '02, vol. 2346, 2002, pp. 83–95.
[91] Y. Wang, D. Chakrabarti, C. Wang, C. Faloutsos, Epidemic spreading in real networks: an eigenvalue viewpoint, in: 22nd Symp. Reliable Distributed Computing, Florence, Italy, 6–8 October, 2003.
[92] T.H. Wei , The Algebraic Foundations of Ranking Theory, Thesis, Cambridge, 1952.
[93] R.C. Wilson, E.R. Hancock, B. Luo, Pattern vectors from algebraic graph theory, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1112–1124.
[94] X. Ying, X. Wu, Randomizing social networks: a spectrum preserving approach, in: Proc. SIAM Internat. Conf. Data Mining, SDM2008, April 24–26, 2008, Atlanta, Georgia, USA, SIAM, 2008, pp. 739–750.