R E S E A R C H   R E P O R T

APPLICATIONS OF GRAPH SPECTRA TO COMPUTER SCIENCE[1]

Graph Spectral Techniques as Tools in Information and Communication Technologies

Dragoš Cvetković, Mathematical Institute SANU, Beograd

with collaborators

Tatjana Aleksić, University of Kragujevac, Faculty of Sciences
Slobodan Simić, Mathematical Institute SANU, Beograd
Dragan Stevanović, University of Niš, Faculty of Sciences
Ljiljana Trajković, Simon Fraser University, Vancouver

Email: ecvetkod@etf.rs , aleksictatjana@yahoo.com , sksimic@mi.sanu.ac.rs ,
dragance106@yahoo.com , ljilja@cs.sfu.ca

March 10, 2010

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

Abstract. *In this report 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.*

Keywords: graph theory, graph spectra, applications, computer science, information technology, communication technology, internet, complex networks

AMS Classification: 05C50 , 68M07, 68P20, 68R10

1

1 Introduction

In this report we shall give a survey of applications of the theory of graph spectra to computer science.

Applications of graph spectra 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 but also to become able to use the literature.

Several papers in computer science cite books on graph spectra such as [Big], [Chu], [CvDSa], [CvRS1]. To document spectral techniques used several books on matrices are cited as well.

In this introductory section we shall present some expository texts on applications of graph spectra.

The books [CvDSa], [CvDGT] contain each a chapter on applications of graph eigenvalues.

The book [CvRS4] 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 [CvGu] 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 [CvGu] 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 [Cve] provides an introduction to the theory of graph spectra and a short survey of applications of graph spectra. There are four sections: 1. Basic notions, 2. Some results, 3. A survey of applications, 4. Selected bibliographies on applications of the theory of graph spectra. 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 and the fifth subsection contains some information about that. In all fields we were forced to give only examples of applications.

2   A survey of applications

It was recognized in about last ten 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 as the titles of the subsections show.

Note that the classification of numerous applications into subsections does not reflect always the importance of the subjects. Also there is an overlapping of the classified material.

We have not included numerous applications of graph spectra to combinatorial optimization although many of the problems of combinatorial optimization appear in computer science. We did so because here we have applications of graph spectra to another branch of mathematics.

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 hypergraphs 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.

Here we mention some general references related to applications of graph spectra in computer science.

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

The paper [Spi] 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.

2.1 Expanders

One of the oldest applications (from 1970's) of graph eigenvalues in Computer Science is related to graphs called *expanders*. Avoiding a formal definition, 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 *superconcentrators*, 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*. For an introduction to this type of applications see [CvSi] and references cited therein. Paper [LuPS] is one of the most important papers concerning Ramanujan graphs.

2.2 Virus propagation in computer networks

The largest eigenvalue λ_1 plays an important role in modelling 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 [WaCWF] that the epidemic threshold in spreading viruses is proportional to $1/\lambda_1$. Motivated by this fact, the authors of [DaKo] determine graphs with minimal λ_1 among graphs with given numbers of vertices and edges, and having a given diameter. Another model of virus propagation in computer networks has been developed in [VaOK] with the same conclusion concerning $1/\lambda_1$.

2.3 Computer vision and pattern recognition

Spectral[2] 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 neighbouring pixels with weight depending on how similar the pixels are.

---

[2]written by Tatjana Aleksić

A more sofisticated idea is to represent an image's content by a graph with specially selected points as vertices. First, a corner detection algorithm has to be used to detect the interest points in the image which could be represented by graph nodes (Harris detector, The Moravec corner detector, etc.). The interest points that these algorithms detect are points in an image which have a well-defined position and can be robustly detected. This means that an interest point can be a corner but it can also be, an isolated point of local intensity maximum or minimum, line endings, or a point on a curve where the curvature is locally maximal. The edges between the neighboring nodes are determined using algorithms such as Delaunay triangulation or, in the case of weighted graphs, the similarity between the nodes.

Several authors have explored the use of the Laplacian and related operators to map data to a manifold in a low dimensional space [1],[2]. Horaud and Sossa [3] 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, Wilson and Hancock [4] 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 [5]. Pattern spaces may then be constructed from the feature vectors using techniques such as principal components analysis.

Principal components analysis - (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. Now it is mostly used as a tool in exploratory data analysis and for making predictive models. PCA involves the calculation of the eigenvalue decomposition of a data covariance matrix usually after mean centering the data for each attribute. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA is theoretically the optimum transform for given data in least square terms. PCA essentially rotates the set of points around their mean in order to align with the first few principal components. This moves as much of the variance as possible (using a linear transformation) into the first few dimensions. The values in the remaining dimensions, therefore, tend to be highly correlated and may be dropped with minimal loss of information. PCA is often used in this manner for dimensionality reduction.

Techniques from spectral-graph theory have been used to develop a powerful array of algorithms in computer vision and pattern recognition. For

5

instance, Shi and Malik [6] 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 so as to maximize the normalized graph cut. Sarkar and Boyer [7] have shown how the eigenvector of the largest eigenvalue of the weighted proximity [3] matrix can be used to group line segments. Random-walk based graph matching methods [8], [9] use the eigenvector of the largest eigenvalue of the adjacency matrix to convert a graph into a string. The attractive feature of eigenvector methods is that they rely on a simple matrix representation of the problem at hand and result in algorithms that do not require complex search procedures or control structures.

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. The idea of graph clustering is to divide graphs into different groups based on the structural properties. Although graph structures have been proved useful in both low-level and high-level vision[4], they are computationally cumbersome because of the need to establish reliable correspondence between nodes. Standard graph clustering methods need to solve the correspondence problems between nodes (of two graphs) first. Again this is a potentially NP-hard problem and the computational overheads can be large. Recently spectral graph theory has been applied to graph clustering [3],[10]. Luo, Wilson and Hancock [4],[5] have proposed spectral invariants for graph clustering. These methods do not need to solve the node correspondence problem, instead they rely on using information from the spectrum of the Laplacian matrix.

The normalized Laplacian spectrum is closely related to the heat kernel of a graph. The heat kernel is the solution of the heat equation and is

---

[3]The global relationship among image features can be very effectively captured in the form of a graph whose nodes represent the image features and whose links 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, and have similar orientation (generalized parallelism). We call this graph the relation graph. In practice, the links in the graph are weighted according to the degree of compatibility between two nodes Our task is to formulate measures which capture global properties of this relation graph. Eigenvalues and eigenvectors of the relation graph (the weighted matrix joined to the relation graph) provide exciting possibilities as a basis for such measures.

Matrica je nazvana proximity matrix zbog rastojanja(proximities between edge segments) koja se u radu koriste za definisanje teina.

[4]Tasks such as edge detection, image segmentation, line detection, motion analysis, etc. are considered to be low-level problems as they detect the constituent parts of objects. These techniques, therefore, determine the features that will be used by high-level vision techniques such as object recognition and scene analysis.

formed by exponentiating the normalized Laplacian eigensystem over time $(h_t = e^{-t\hat{L}})$. The heat kernel matrix of the graph encapsulates the way in which information flows through the edges of the graph over time. Methods to extract useful and stable invariants from the heat kernel, as a means of graph clustering, have been explored. They have been also used to map the nodes of a graph to points in a vector space [11]. This is achieved by the analysis of the heat kernel.

The trace of the heat kernel [12],[13] be used for the purpose of characterizing graphs. The trace of the heat kernel is found by summing a series of terms, each of which is the result of exponentiating a normalized Laplacian eigenvalue with time $(Tr[t] = \sum_{i=1}^{|V|} e^{-\mu_i t})$. As a result the heat kernel trace is a function whose parameters are the normalized Laplacian eigenvalues and whose argument is time. The shape of this function can be used to characterize the corresponding graph [14].

The heat content is defined as the sum of the elements of the heat kernel and can be expanded as a polynomial over time $(Q(t) = \sum_{u \in V} \sum_{v \in V} h_t(u,v))$.

The coefficients of the polynomial are known to be invariants. It has been demonstrated how the polynomial coefficients can be computed from the normalized Laplacian eigensystem [15], [16]. Graph clustering is performed by applying principal components analysis to vectors constructed from the polynomial coefficients. It has been shown that manifold learning theory and spectral methods can be combined to solve the image classification problem.

There has recently been an increasing interest in hypergraph-based methods for representing and processing visual information extracted from images. The main reason for this is that hypergraph representations allow nodes to be multiply connected by edges, and can hence capture multiple relationships between features. The idea has been to extend techniques from spectral-graph theory to hypergraphs. A hypergraph model for characterizing object structures has been established and the spectral method to construct pattern vectors from the hypergraph's Laplacian and characteristic polynomials has been used [17], [18]. The authors apply feature vectors to clustering hypergraphs extracted from images of different object views and demonstrate their effectiveness in hypergraph characterization. Hypergraph-based spectral methods, however are relatively new concepts and there is more to be explored.

2.4 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, 44(2003), No. 3.

It has been discovered recently [ChDEL] that integral graphs can play a role in the so called perfect state transfer in quantum spin networks. Further details on this topic can be found in [SaSS].

2.5 Load balancing in multiprocessor systems

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 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 cancelled 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.

In this way we have defined a load flow on the edge set of G. First, particular amounts of load flow should be considered algebraically, i.e., having in mind their sign. So, if $x_i^{(k-1)}$ is negative, then vertex i, in fact, receives the corresponding amount. For each edge ij we have two parts of the flow: the part which is sent (or received) by i and the part which is sent (or received) by j. These two amounts should be added algebraically and in this way we get final value of the flow through edge ij. This flow at the end has a non-negative value which is sent either from i to j or vice versa.

The number of iterations in (1) is equal to the number of non-zero distinct Laplacian eigenvalues of the underlying graph. Hence we see that from

9

the point of view of complexity of the load balancing algorithms graphs with a small number of distinct Laplacian eigenvalues are suitable for modelling multiprocessor interconnection networks. 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 [ElKM] as a parameter relevant for the choice and the design of multiprocessor interconnection networks.

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 [CvDa2]. The further study of integral graphs in connection to multiprocessor topologies seems to be a promising subject for future research.

See references [DeMP, DiFM, GhMS, HuBl, HuBE] for further information on the load balancing problem.

2.6   Multiprocessor interconnection networks

As we have already pointed out in the previous subsection, 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 [ElKM] related to the design of multiprocessor topologies. The main conclusion of [ElKM] 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 [CvDa1, CvDa2, CvDa3].

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 Complex networks

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 show 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 low* of the form $x^{-\beta}$ for a positive 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 [FaFF] 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 [MiPa].

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 $\lambda_i = ai^{-\gamma}$ for some constant a and positive γ.

The book [ChLu] is devoted to complex networks. There are two chapters which describe spectral properties of such networks.

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

2.8   Internet topology

Studying[5] and modelling 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.

Analyzing the Internet topology using randomly generated graphs, where routers are represented by vertices and transmission lines by edges, has been widely replaced by mining data that capture information about Internet Autonomous Systems and by exploring properties of associated graphs on the AS-level. The Route Views data [6] and RIPE [7] datasets collected from Border Gateway Protocols (BGP) routing tables have been extensively used by the research community [FaFF, SiFFF, ChCGJSW]. The discovery of power-laws and spectral properties of the Internet topology indicates a complex underlying network infrastructure.

Analysis of the collected datasets indicates that the Internet topology is characterized by the presence of various power-laws observed when considering a node degree vs. node rank, a node degree frequency vs. degree, and a number of nodes within a number of hops vs. number of hops [FaFF, SiFFF]. Some of these early conclusions were subsequently revised by considering a more complete AS-level representation of the Internet topology [ChCGJSW, ChGJSW]. These extended maps 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 [GkMZ, NaST]. Power-laws also appear in the eigenvalues of the adjacency matrix and the normalized Laplacian matrix vs. the order of the eigenvalues. They also show invariance regardless of the exponential growth of the Internet.

While various power-law exponents associated with the Internet topology have remained similar over the years, indicating that the power-laws do not capture every property of a graph and are only one measure used

---

[5]written by Ljiljana Trajković

[6](2010, February 10). BGP capture datasets [Online]. Available: http://archive.routeviews.org.

[7](2010, February 10). RÄ©seaux IP EuropÄ©ens [Online]. Available: http://www.ripe.net/ris.

to characterize the Internet, spectral analysis of both the adjacency matrix and the normalized Laplacian matrix of the associated graphs reveals new historical trends in the clustering of AS nodes and their connectivity. The eigenvectors corresponding to the largest eigenvalues of the normalized Laplacian matrix have been used to identify clusters of AS nodes with certain characteristics [GkMZ]. Spectral analysis was employed to analyze the Route Views and RIPE datasets in order to find distinct clustering features of the Internet AS nodes [ChTr]. For example, the connectivity graphs of these datasets indicate visible changes in the clustering of AS nodes and the AS connectivity over the period of five years [NaST]. Clusters of AS nodes can be also identified based on the eigenvectors corresponding to the second smallest and the largest eigenvalue of the adjacency matrix and the normalized Laplacian matrix [SuTr]. The connectivity and clustering properties of the Internet topology can be further analyzed by examining element values of the corresponding eigenvectors.

2.9 Internet search

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

The structure of the Internet is represented by a digraph G where web pages correspond to vertices and links to arcs.

HITS exploits eigenvectors belonging to the largest eigenvalues of the matrixces AA^T and $A^T A$ 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 defines 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 matrix is a transition matrix of a Markov chain and the normalized eigenvector of the largest eigenvalue of its transpose defines the stationary state of the chain. Pages are ranked by the coordinates of this eigenvector.

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

2.10 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 [Saw], [Ski].

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 [GkMZ] spectral filtering is applied in the study of Internet structure.

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

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 [ShDSZ].

2.11 Statistical databases

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 parameters, 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 *secure*.

One can consider a restricted case where the query collection can be described as a graph. Surprisingly, the results from [Bra, BrMS] show an amazing connection between compromise-free query collections and graphs with least eigenvalue -2. This connection was recognized in the paper [BraCv].

It is interesting to note that original Doob's description [Doo] 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, Doob and Simić [CvDS] in 1981 in terms of the chain groups, not explicitly dealing with cycles and dumbbells. The independent discovery of Branković, Miller and Širáň [BrMS] in 1996 put implicitly some light on the description of the eigenspace in generalized line graphs a bit before Cvetković, Rowlinson and Simić in 2001 (the paper [CvRS2] was submitted in 1998), using the star complement technique and without being aware of [BrMS], gave the entire description of the eigenspace.

2.12   Social networks

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) [YiWu].

References

[Big] Biggs N.L., *Algebraic Graph Theory*, Cambridge University Press, Cambridge,

[Bra] Branković L., *Usability of secure statistical data bases*, PhD Thesis, Newcastle, Australia, 1998.

[BraCv] Branković Lj., Cvetković D., *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.

[BrMS] Branković L., Miller M., Širáň J., *Graphs, (0,1)-matrices and usability of statistical data bases*, Congressus Numerantium, 120 (1996), 186–192.

[BrPa] Brin S., Page L., *The Anatomy of Large-Scale Hypertextual Web Search Engine*, Proc. 7th International WWW Conference, 1998.

[ChGJSW] H. Chang, R. Govindan, S. Jamin, S. Shenker, W. Willinger, Towards capturing representative AS-level Internet topologies, in *Proc. ACM SIGMETRICS Performance Evaluation Review*, New York, NY, June 2002, vol. 30, no. 1, pp. 280–281.

[ChCGJSW] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, W. Willinger, The origin of power laws in Internet topologies revisited, in *Proc. INFOCOM*, New York, NY, USA, Apr. 2002, pp. 608–617.

[ChTr] Chen, J., Trajković L., *Analysis of Internet topology data*, Proc. IEEE Internat. Symp., Circuits and Systems, ISCAS 2004, Vancouver, B.C., May 2004, 629–632.

[ChDEL] Christandl M., Datta N., Ekert A., Landahl A.J., *Perfect state transfer in quantum spin networks*, Phys. Rev. Lett., 92(2004),187902.

[Chu] Chung F., *Spectral Graph tTheory*, American Mathematical Society, Providence, Rhode Island, 1997.

[ChLu] Chung F., Lu L., *Complex Graphs and Networks*, American Mathematical Society, Providence, Rhode Island, 2006.

[Cve] Cvetković D., *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, 7-31.

[CvDa1] Cvetković D., Davidović D., *Applications of some graph invariants to the analysis of multiprocessor topologies*, YUJOR, 18(2008), No. 2, 173-186.

[CvDa2] Cvetković D., Davidović T., *Multiprocessor interconnection networks with small tightness*, Internat. J. Foundations Computer Sci., 20(2009), No. 5, 941-963. doi: 10.1142/S0129054109006978

[CvDa3] 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.

[CvDGT] Cvetković D., Doob M., Gutman I., Torgašev A., *Recent Results in the Theory of Graph Spectra*, North-Holland, Amsterdam, 1988.

[CvDGT] Cvetković D., Doob M., Gutman I., Torgašev A., *Recent Results in the Theory of Graph Spectra*, North-Holland, Amsterdam, 1988.

[CvDSa] Cvetković D., Doob M., Sachs H., *Spectra of Graphs, Theory and Application*, 3rd edition, Johann Ambrosius Barth Verlag, Heidelberg–Leipzig, 1995.

[CvDS] Cvetković D., Doob M., Simić S., *Generalized Line Graphs*, J. Graph Theory, 5 (1981), No.4, 385–399.

[CvGu] Cvetković D., Gutman I., ed., *Applications of Graph Spectra*, Zbornik radova 13(21), Mathematical Institute SANU, Belgrade, 2009.

[CvRo] Cvetković D., Rowlinson P., *The largest eigenvalue of a graph – a survey*, Lin. Multilin. Algebra, 28 (1990), 3–33.

[CvRS1] Cvetković D., Rowlinson P., Simić S. K., *Eigenspaces of Graphs*, Cambridge University Press, Cambridge, 1997.

[CvRS2] Cvetković D., Rowlinson P., Simić S. K., *Graphs with least eigenvalue -2: the star complement technique*, J. Algebraic Comb., 14 (2001), 5–16.

[CvRS3] Cvetković D., Rowlinson P., Simić S. K., *Spectral Generalizations of Line Graphs, On Graphs with Least Eigenvalue* -2, Cambridge University Press, Cambridge, 2004.

[CvRS4] Cvetković D., Rowlinson P., Simić S. K., *An Introduction to the Theory of Graph Spectra*, Cambridge University Press, Cambridge, 2009.

[CvSi] 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.

[DaKo] van Dam E. R., Kooij R. E., *The minimal spectral radius of graphs with a given diameter*, Linear Alg. Appl. 423 (2007), 408–419.

[DeMP] T. Decker, B. Monien, R. Preis, *Towards optimal load balancing topologies*, in *Proc. 6th EuroPar Conf.*, Lect. Notes Comput. Sci. 1900, Springer, Berlin, p.p. 277–287

[DiFM] R. Dickmann, A. Frommer, B. Monien, *Efficient scheme for nearest neighbor load balancing*, Parallel Comput. 25:789–812, 1999.

[Doo] Doob M., *An interrelation between line graphs, eigenvalues, and matroids*, J. Combinatorial Theory, Ser. B, 15 (1973), 40–50.

[ElKM] R. Elsässer, R. Královič, B. Monien, *Sparse topologies with small spectrum size*, Theor. Comput. Sci. 307:549–565, 2003.

[FaFF] 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.

[GhMS] B. Ghosh, S. Muthukrishnan, M. H. Schultz, *First and second order diffusive methods for rapid, coarse, distributed load balancing*, in: *Proc. 8th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'96)*, ACM Press, 1996, pages 72–81.

[GkMZ] Gkantsisis C., Mihail M., Zegura E., *Spectral analysis of internet topologies*,Proc. INFOCOM2003, San Francisco, CA, Vol. 1, march 2003, 364-374.

[HuBl] Y. F. Hu, R. J. Blake, *An improved diffusion algorithm for dymanic load balancing*, Parallel Comput. 25:417–444, 1999.

[HuBE] Y. F. Hu, R. J. Blake, D. R. Emerson, *An optimal migration algorithm for dymanic load balancing*, Concurrency Pract. Exp. 10:467–483, 1998.

[Kle] Kleinberg J., *Authoratitive sources in a hyperlinked environment*, J. ACM, 48 (1999), 604-632.

[LaMe] Langville A.N., Meyer C.D., *A survey of eigenvector methods for Web information retrivial*, SIAM Rev., 47(2005), No. 1, 135-161.

[LiWM] Liu D., Wang H., Van Mieghem P., *Spectral perturbation and reconstructability of complex networks*, Phys. Rev. E, 81(2010), 016101, 1-9.

[LuPS] Lubotzky A., Phillips R., Sarnak P., *Ramanujan graphs*, Combinatorica, 8 (1988), 261-277.

[Lux] Luxburg U. von, *A tutorial on spectral clustering*, Stat. Comput. 17(2007), 395-416.

[MiPa] Mihail M. Papadimitrou C.H., *On the eigenvalue power-low*, RANDOM 2002, LNCS 2483, Springer, Berlin, 2002, 254-262.

[NaST] M. Najiminaini, L. Subedi, Lj. Trajković, Analysis of Internet topologies: a historical view, *Proc. IEEE Int. Symp. Circuits and Systems*, Taipei, Taiwan, May 2009, pp. 1697–1700.

[Saw] 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.

[SaSS] Saxena, N., Severini, S., Shparlinski, I. E., *Parameters of integral circulant graphs and periodic quantum dynamics*, Intern. J. Quantum Inf., 5 (2007), No. 3, 417–430.

[ShDSZ] Shokoufandeh A., Dickinson, S. J., Siddiqi K., Zucker S. W., *Indexing using a spectral encoding of topological structure*, IEEE Trans. Comput. Vision Pattern Recognition, 2 (1999), 491–497.

[SiFFF] G. Siganos, M. Faloutsos, P. Faloutsos, C. Faloutsos, Power-laws and the AS-level Internet topology, *IEEE/ACM Trans. Networking*, vol. 11, no. 4, pp. 514–524, Aug. 2003.

[Ski] Skilicorn D.B., *Understanding Complex Data Bases: Data Mining with Matrix Decompositions*, Chapman & Hall/CRC, New York, 2007.

[Spi] Spielman D. A., *Spectral Graph Theory and its Applications*, 48th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2007, 29–38.

[SuTr] L. Subedi, Lj. Trajković, Spectral analysis of Internet topology graphs, to be presented at *IEEE Int. Symp. Circuits and Systems*, Paris, France, May 2010.

[Van] Van Mieghem P., *Performance Analysis of Communications Networks and Systems*, Cambridge University Press, Cambridge, 2006.

[VaOK] Van Mieghem P., Omić J., Kooij R., *Virus spread in networks*, preprint.

[WaCWF] Wang Y., Chakrabarti D., Wang C., Faloutsos C., *Epidemic spreading in real networks: An eigenvalue viewpoint*, 22nd Symp. Reliable Distributed Computing, Florence, Italy, Oct. 6–8, 2003.

[YiWu] Ying X., 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.

References used in Subsection 2.3

1. Roweis S., Saul L. (2000). Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500):2323-2326.

2. Belkin M., Niyogi P. (2000). Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation, 15:1373-1396.

3. Horaud R., Sossa H. (1995). Polyhedral object recognition by indexing. Pattern Recognition, 28(12):1855-1870.

4. Luo B., Wilson R.C., Hancock E. R. (2003). Spectral embedding of graphs. Pattern Recognition, 36:2213-2230.

5. Wilson R.C., Luo B., Hancock, E.R. (2003). Pattern vectors from algebraic graph theory. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27:2220-2237.

6. Shi J., Malik J. (1997). Normalized cuts and image segmentation. International Conference on Computer Vision and Pattern Recognition.

7. Sarkar S., Boyer K. L. (1998). Quantitative measures of change based on feature organization:Eigenvalues and eigenvectors. Computer Vision and Image Understanding, 71(1):110-136.

8. Robles-Kelly A., Hancock E. R. (2005a). Graph edit distance from spectral seriation. IEEE TPAMI, 27:365-378.

9. Caelli T., Kosinov S. (2004). An eigenspace projection clustering method for inexact graph matching. IEEE Trans. Pattern Anal. Mach. Intell., 26(4):515-519.

10. Shokoufandeh A., Dickinson S., Siddiqi K., Zucker S. (1999). Indexing using a spectral encoding of topological structure. International Conference on Computer Vision and Pattern Recognition.

11. Xiao Bai, Edwin Hancock, "Recent Results on Heat Kernel Embedding of Graphs" 5th Graph Based Representation on Pattern Recognition, France, pages 373-382 , 2005

12. Yau S. T., Schoen R. M. (1988). Differential geometry. Science Publication.

13. Chung F. R. K., Yau S. T. (1997). A combinatorial trace formula. Tsinghua lectures on geometry and analysis.

14. Xiao Bai, Edwin Hancock "Trace Formula Analysis of Graphs" Syntactical and Structural Pattern Recognition (SSPR), Portugal, pages 306-313 , 2006

15. Xiao Bai, Edwin Hancock "Graph Clustering using Heat Content Invariants" 2nd Iberian Conference on Pattern Recognition and Image Analysis, Portugal, pages 123-130, 2005

16. Xiao Bai, Edwin Hancock "Clustering Shapes using Heat Content Invariants" The International Conference on Image Processing(ICIP), 2005

17. Peng Ren, Tatjana Aleksić, Richard C. Wilson, Edwin R. Hancock: Hypergraphs, Characteristic Polynomials and the Ihara Zeta Function. CAIP 2009: 369-376

18. Peng Ren, R. C. Wilson, Edwin R. Hancock. Spectral Embedding of Feature Hypergraphs. SSPR 2008, Lecture Notes in Computer Science, 5342, pp. 308-317, 2008.