SPECTRAL RECOGNITION OF GRAPHS

A Survey of Applications in Computer Science

Dragoš Cvetković

Mathematical Institute SANU,
P.O. Box 367, 11000 Belgrade, Serbia
E-mail: ecvetkod@etf.rs
http://www.mi.sanu.ac.rs/cv/cvcvetkovic.htm
(ecvetkod@etf.rs)
Presented at the Conference on Applications of Graph spectra in Computer Science,
Barcelona, July 16 to 20, 2012

July, 2012

Content

Introduction
Preliminaries: Spectral graph theory in computer sciences
Preliminaries: Spectral recognition problems
Graph spectra in mathematics and computer science
Spectral characterizations of graphs
Similarity of graphs
Spectral distances in graphs
Refinements using graph angles
Network alignment problem
Queries for databases and the subgraph isomorphism problem
Structural and spectral perturbations of graphs
Some other spectral recognition problems

Spectral graph theory in computer sciences

Cvetković D., Simić S.K., *Graph spectra in computer science*, Linear Algebra Appl., 434(2011), 1545-1562.

Arsić B., Cvetković D., Simić S.K., Škarić M., *Graph spectral techniques in computer sciences*, Applicable Analysis and Discrete Mathematics, 6(2012), No. 1, 1-30.

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

The second paper contains, among others, the following sections and
subsections:
3. Significant eigenvalues, 3.1. Largest eigenvalue, 3.2. Algebraic
connectivity, 3.3. The second largest eigenvalue, 3.4. The least eigenvalue,
3.5. Main eigenvalues,
4. Eigenvector techniques, 4.1. Principal eigenvector, 4.2. The Fiedler
eigenvector, 4.3. Other eigenvectors,
5. Spectral recognition problems, 6. Spectra of random graphs, 7.
Miscellaneous topics, 7.1. The Hoffman polynomial, 7.2. Integral graphs,
7.3. Graph divisors.

Spectral recognition problems

The whole spectral graph theory is related in some sense to the recognition of graphs since spectral graph parameters contain a lot of information on the graph structure (both global and local). However, we shall treat here the problems of recognizing entire graphs, or some parts of them, both in an exact manner and in an approximative way.
In particular, we shall consider
- characterizations of graphs with a given spectrum
- exact or approximate constructions of graphs with a given spectrum,
- similarity of graphs,
- perturbations of graphs.

Graph spectra in mathematics and computer science

Spectral graph theory is a very well developed mathematical field but also an engineering discipline.

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

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

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

Moreover, in some applications in data mining graph spectra are used to encode graphs themselves.

The following example is illustrative in this respect. The indexing structure of objects appearing in computer vision (and in 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

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.

In some cases researchers feel that the spectrum very well characterizes the graphs under consideration so that the spectrum is considered as a *fingerprint* of the corresponding network. The eigenvalues γ_i; $i = 1, 2, \ldots, n$ of the normalized Laplacian $\hat{L}$ in non-decreasing order can be represented by points $(\frac{i-1}{n-1}, \gamma_i)$ in the region $[0, 1] \times [0, 2]$ and can be approximated by a continuous curve. It was noticed in

Vukadinović D., Huang P., Erlebach T., *On the spectrum and structure of the Internet topology graphs*, Proc. Second Internat. Workshop on Innovative Internet Computing Systems, IICS '02, vol. 2346, 2002, 83-95.

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.

Spectral characterizations of graphs

At some time it was conjectured that non-isomorphic graphs have different spectra, i.e. that graphs are characterized by their spectra. Very quickly this conjecture was refuted and numerous examples and families of non-isomorphic graphs with the same spectrum were found. In particular, it was proved that almost all trees are not characterized by their spectra. Analogous question for general graphs remained open.

Also in Chemistry there was a criticism on using graph eigenvalues to characterize molecules
Heilbronner E., Jones T.B., *Spectral differences between "isospectral" molecules*, J. Amer. Chem. Soc., 100(1978), No. 20, 6505-6507.

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

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

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

The paper
Haemers W., Spence E., *Enumeration of cospectral graphs*, European J.
Combin. 25 (2004), 199-211.
provides spectral uncertainties r_n with respect to the adjacency matrix A,
s_n with respect to the Laplacian L and q_n with respect to the signless
Laplacian Q of sets of all graphs on n vertices for $n \leq 11$:

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

We see that the sequences s_n and q_n are decreasing for $n \leq 12$ while the sequence r_n is increasing for $n \leq 10$. Yet, it starts to decrease for $n > 10$.

This is a strong basis for believing that almost all graphs are determined by their spectra when n tends towards the infinity, as conjectured in Dam E.R. van, Haemers W., *Which graphs are determined by their spectrum?*, Linear Algebra Appl., 373(2003), 241-272.

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

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

Recently, a spectral theory of graphs based on the signless Laplacian has been developed

Cvetković D., Simić S.K., *Towards a spectral theory of graphs based on the signless Laplacian, III*, Appl. Anal. Discrete Math., 4(2010), 156-166.

There are many results in the mathematical literature on spectral characterizations of particular classes of graphs. For example, complete graphs, paths and circuits are characterized by their A-spectra up to an isomorphism.

Graphs characterized by their spectra up to an isomorphism are also called in DS-graphs or it is said that such graphs are DS (spectrally determined).

There are also characterizations with some exceptional cospectral mates.

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

Similarity of graphs

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

Spectral distances in graphs

The Euclidean distance between the eigenvalue sequences of two graphs on the same number of vertices is called the *spectral distance* of graphs. Some other spectral distances have been considered as well (e.g., the Manhattan distance).
In defining spectral distances various graph matrices can be used (e.g., the adjacency matrix, the Laplacian, the signless Laplacian).

Some mathematical results on the Manhattan spectral distance have been obtained in

Jovanović I. Stanić Z., *Spectral distances of graphs*, Linear Algebra Appl. (2011), doi:10.1016/j.laa.2011.08.019

An interesting observation from this paper is that the Manhattan distance is in connection with graph energy, a graph invariant studied very much in the literature. The *energy* of a graph is the sum of absolute values of its A-eigenvalues. Thus the energy of a graph is the Manhattan spectral distance of the graph from the graph without edges.

The use of the Laplacian and the signless Laplacian matrix for the Manhattan distance seems to be very useful when considering subgraphs.

By the interlacing theorems for these matrices, all eigenvalues go down or remain the same if an edge is deleted from the graph. Hence the distance between a graph and any of its edge deleted subgraphs is equal to the decrement of the trace of the matrix. Since for both matrices the trace is equal to the sum of vertex degrees, we conclude that the distance is equal to the twofold number of deleted edges.

All these properties do not hold for the adjacency matrix.

Two graphs are considered as *similar* if their spectral distance is small. If two graphs are at zero distance, this does not necessarily mean that they are equal (i.e. isomorphic); they are only cospectral. In this sense, cospectral graphs are similar.

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

An algorithm for such a spectral graph reconstruction has been developed in

Comellas F., Diaz-Lopez J., *Spectral reconstruction of complex networks*, Phys. A, 387(2008), 6436-6442.

Given the spectrum of a graph, the algorithm starts from a random graph and uses the tabu search to diminish the Euclidean spectral distance between the given and the current spectrum.

Both, the distance and the meta-heuristic, can be varied. We could use the Manhattan distance based on the adjacency matrix or on the signless Laplacian. The tabu search could be replaced by the variable neighbourhood search or by some other meta-heuristics.
Computer programs for spectral reconstruction of graphs can be used to generate example graphs with desired spectral properties.

The variable neighbourhood search is exploited in the programming package AutoGraphiX (briefly AGX) for finding graphs with extremal values of a graph invariant chosen by the user. The system starts from a random graph or from a graph given by the user. The graph is perturbed to some extent using the variable neighbourhood search and a new graph is chosen which improves maximally the considered graph invariant. The system AGX is very useful in formulating some conjectures which are later treated by theoretical means.

For example, system AGX has generated several conjectures for the energy of a graph

Caporossi G., Cvetković D., Gutman I., Hansen P., *Variable neighborhood search for extremal graphs*, 2. *Finding graphs with extremal energy*, J. Chem. Inform. Comp. Sci., **39**(1999),984-996.

and thirty conjectures concerning signess Laplacian eigenvalues

Cvetković D., Rowlinson P., Simić S., *Eigenvalue bounds for the signless Laplacian*, Publ. Inst. Math. (Beograd), 81(95)(2007), 11-27.

It would be interesting to treat some conjectures from
Jovanović I. Stanić Z., *Spectral distances of graphs*, Linear Algebra Appl.
(2011), doi:10.1016/j.laa.2011.08.019
concerning spectral distances of graphs by AGX (for example, conjectures
on graphs with maximal spectral distance).

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

Refinements using graph angles

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

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

Let G be a graph on n vertices with distinct eigenvalues $\mu_1, \mu_2, \ldots, \mu_m$ ($\mu_1 > \mu_2 > \cdots > \mu_m$) and let $S_1, S_2, \ldots, S_m$ be the corresponding eigenspaces. Let $\{e_1, e_2, \ldots, e_n\}$ be the standard (orthonormal) basis of R^n. The numbers $\alpha_{pq} = \cos\beta_{pq}(p = 1, 2, \ldots, m; q = 1, 2, \ldots, n)$, where β_{pq} is the angle between S_p and e_q, are called *graph angles*.

The sequence α_{pq} ($q = 1, 2, \ldots, n$) is called the *eigenvalue angle sequence* corresponding to the eigenvalue μ_p ($p = 1, 2, \ldots, m$).

We also define the *angle matrix* of G, i.e. an $m \times n$ matrix (m is the number of its distinct eigenvalues, while n is the order of G) as a matrix (α_{ij}). This matrix is a graph invariant if its columns are ordered lexicographically.

The rows of the angle matrix are called the *standard eigenvalue angle sequences*.

Let $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ $(i = 1, 2, \ldots, n)$ be orthonormal eigenvectors of G. Define $M_p = \{j \mid Ax_j = \mu_p x_j\}$. We have $\alpha_{pq}^2 = \sum\limits_{j \in M_p} x_{jq}^2$ for squares of angles of G.

An overview of results on graph angles is given in
Cvetković D., Rowlinson P., Simić S. K., *Eigenspaces of Graphs*,
Cambridge University Press, Cambridge, 1997.
including the characterizing properties of graph angles.

It was suggested in

Cvetković D., *Characterizing properties of some graph invariants related to electron charges in the Hückel molecular orbital theory*, Proc. DIMACS Workshop on Discrete Mathematical Chemistry, DIMACS Ser. Discrete Math. Theoret. Comp, Sci., **51**(2000), 79-84.

that cospectral graphs can be ordered by graph angles, in particular, lexicographically by their standard eigenvalue angle sequences.

The paper provides an example of 21 mutually cospectral graphs (on 10 vertices with 20 edges) ordered by the first standard eigenvalue angle sequences.

In defining a spectral graph distance we use differences between corresponding eigenvalues of two graphs.

For a set of cospectral graphs and for each spectral graph distance we can define the corresponding *cospectral graph distance* by using differences between the corresponding entries of the angle matrix instead of differences between corresponding eigenvalues.

For example, the *Manhattan cospectral graph distance* is the sum of absolute values of differences between the corresponding entries of the angle matrices of the graphs.

However, the cospectral graph distance can be 0 for non-isomorphic graphs since there are graphs having the same eigenvalues and the same angles (for example, strongly regular graphs, see Cvetković D., Rowlinson P., Simić S. K., *Eigenspaces of Graphs*, Cambridge University Press, Cambridge, 1997.

Such effects cannot appear if we use stronger invariants such as *canonical star basis* of a graph. However, the computational complexity of constructing a canonical star basis is probably so high that it is not practical to use this approach in applications.

Network alignment problem

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

Given the graphs G and H, a measure of similarity between vertices of G and vertices of H is introduced by some definition. Let $R_{i,j}$ be the measure of similarity between vertex i of G and vertex j of H. Let B be the bipartite graph on vertex sets of the graphs G and H with edge weights $R_{i,j}$.

We want to find a maximal matching with a maximal sum of weights in B. This matching defines subgraphs of graphs G and H which are similar w.r.t. the introduced similarity between vertices of G and vertices of H. Finding a maximal matching with a maximal sum of weights in a bipartite graph can efficiently be performed by existing algorithms for the assignment problem in combinatorial optimization. See, for example, Papadimitriou C., Steiglitz K. *Combinatorial optimization: algorithms and complexity*, Dover, 1998.

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

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

A survey of structural (non-spectral) measures of similarity between vertices can be found in

Xuan Q., Yu L., Du F., Wu T.-J., *A review on node-matching between networks*, New Frontiers in Graph Theory, Ed. Y. Zhang, Intex, Rijeka, 2012, 153-167.

The normalized positive eigenvector belonging to the largest A-eigenvalue of a connected graph is called the *principal eigenvector*.
The principal eigenvector of a graph can efficiently be computed by an iterative algorithm called the *power method* (see, for example, Golub G.H., Van Loan C. *Matrix computations*, Johns Hopkins University Press, 2006.).

A spectrally based measure of similarity between vertices of two graphs is used in

Singh R., Xu J., Berger B., *Pairwise global alignment of protein interaction networks by matching neighborhood topology*, Research in Computational Molecular Biology, Springer, 2007, 16-31.

in the context of protein interaction networks. The algorithm is called IsoRank, it uses graph eigenvectors and is similar to the algorithm PageRank, used in the Internet search

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

Here the measure of similarity $R_{i,j}$ is equal to the product of the i-th coordinate of the principal eigenvector of the graph G and the j-th coordinate of the principal eigenvector of the graph H. This measure has the property that for any pair (i, j) it is equal to the mean value of similarities between all pairs (p, q) where p ia a neighbour of i and q ia a neighbour of j.

The paper
Singh R., Xu J., Berger B., *Pairwise global alignment of protein interaction networks by matching neighborhood topology*, Research in Computational Molecular Biology, Springer, 2007, 16-31.
uses graph product $G \times H$ of graphs G and H. The quantity $R_{i,j}$ is interpreted as a coordinate of the principal eigenvector of $G \times H$. However, it is well-known that the adjacency matrix of $G \times H$ is equal to the Kronecker product of adjacency matrices of graphs G and H and that the principal eigenvector of $G \times H$ is the Kronecker product of principal eigenvectors of graphs G and H. Hence, principal eigenvectors of graphs G and H can be computed separately.

A similar concept of the measure of similarity has been introduced in Blondel V.D., Gajardo A., Heymans M., Senellart P., Van Doren P., *A measure of similarity between graph vertices: applications to synonym extraction and web searching*, SIAM Rev., 46(2004), No. 4, 647-666. even more generally between vertices of two digraphs and applied to synonym extraction from a dictionary. When considering undirected graphs, we obtain the construction used in IsoRank.

Coordinates of the principal eigenvector are related to vertex neighbourhoods because they are asymptotically proportional to the number of walks of length k starting at particular vertices. The following relevant theorem of T.H. Wei stems from

Wei T.H., The algebraic foundations of ranking theory, Thesis, Cambridge, 1952.

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 principal eigenvector of G.*

Counting walks with specified properties in a graph (or digraph) is related to graph spectra by the following well-known result.

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

The paper
Kuchaiev O., Milenković T., Memišević V., Hayes W., Pržulj N.,
Topological network alignment uncovers biological function and phylogeny,
J. Roy. Soc. Interface,7(2010), 1341-1354.
surveys methods of network alignment used in protein interaction networks
and recommends an algorithm based on "graphlets" (induced subgraphs
on at most 5 vertices). This is again a non-spectral approach in which a
vertex is characterized by a 73-dimensional vector (vertex signature) whose
coordinates represent frequencies of graphlets in which the vertex appears.
The corresponding programming package is called GRAAL.

Approach using graph angles

Let G be a graph with adjacency matrix A, and let $N_k(j) = a_{jj}^{(k)}$, the number of walks of length k in G originating and terminating at vertex j. Let $\mathcal{H}_j(t)$ be the generating function $\sum_{k=0}^{\infty} N_k(j) t^k$. We can obtain

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

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

We think that in problems of network alignment vertices should be characterized by generating functions $\mathcal{H}_j(t)$. This function depends on the vertex neighbourhood which is in this case extended to the whole graph unlike the method with graphlets where the neighbourhood is very limited. For example, the measure of similarity $R_{i,j}$ can be defined in some way using the difference $\mathcal{H}_i^G(t) - \mathcal{H}_j^H(t)$ of generating functions of vertex i in G and vertex j in H.

The following formulas are also useful.

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

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

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

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

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

since

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

It would be interesting to compare the performances of the three described approaches to the network alignment problem using various criteria. It was reported in

Kuchaiev O., Milenković T., Memišević V., Hayes W., Pržulj N., *Topological network alignment uncovers biological function and phylogeny*, J. Roy. Soc. Interface,7(2010), 1341-1354.

that GRAAL finds common subgraphs with more vertices than IsoRank

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

Queries for databases and the subgraph isomorphism problem

In several databases the data are often represented as graphs. Very frequently graphs are indexed by their spectra.
In
Pinto A., *MIREX2007 - Graph spectral method*, to appear.
a spectral graph theory approach is presented for representing melodies as graphs, based on intervals between the notes they are composed of. These graphs are then indexed using their Laplacian spectrum. This makes it possible to find melodies similar to a given melody.
The query for such a database is given by a graph. To find similar data in the database it is necessary to compare subgraphs of the query graph with subgraphs of the graphs stored in the database. One should efficiently select a small set of database graphs, which share a subgraph with the query.

The *subgraph isomorphism problem* is a computational task in which two graphs G and H are given as input, and one must determine whether G contains a subgraph that is isomorphic to H. Sometimes the name *subgraph matching* is also used for the same problem. Subgraph isomorphism is a generalization of two well-known NP-complete problems, the maximum clique problem and the problem of testing whether a graph contains a Hamiltonian cycle, and is therefore NP-complete.

There is a similar problem known as the *maximum common subgraph-isomorphism problem*. This problem is known to be NP-hard. The formal description of the problem is as follows: given two graphs G_1 and G_2, what is the largest induced subgraph of G_1 isomorphic to an induced subgraph of G_2? The associated decision problem, i.e., given G_1, G_2 and an integer k, deciding whether G_1 contains an induced subgraph of at least k edges isomorphic to an induced subgraph of G_2 is NP-complete.

Instead of comparing subgraphs one can compare their spectra. While the subgraph isomorphism problem is NP-complete, comparing spectra can be done in polynomial time.

The so called Interlacing Theorem plays an important role in problems of spectral graph recognition and in spectral graph theory and its applications in general.

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

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

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

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

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

Theorem. *Let G be a connected graph on n vertices. Eigenvalues in non-decreasing order of the Laplacian $L = D - A$ of G are denoted by $\nu_1 = 0, \nu_2, \ldots, \nu_n$. Let G' be obtained from G by adding an edge and let $\sigma_1 = 0, \sigma_2, \ldots, \sigma_n$ be L-eigenvalues of G'. Then*

$$0 = \nu_1 = \sigma_1 \leq \nu_2 \leq \sigma_2 \leq \cdots \leq \nu_n \leq \sigma_n.$$

The proof is obtained using well-known Courant-Weyl inequalities.

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

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

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

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.

Using this fact a spectral coding of graphs is introduced in
Zou L., Chen L., Yu J.X., Lu Y., *A novel spectral coding in a large graph
database*, EDBT'08, March 25 - 30, 2008, Nantes, France.
This coding uses spectra of trees associated to graph vertices.

It seems that better search pruning possibilities can be expected from
interlacing theorems for Laplacian and signless Laplacian spectra.
Laplacian eigenvalues are used to code graphs, for example, in
Demirci M.F., Leuken R.H. van, Veltkamp R.C., *Indexing through
laplacian spectra*, Computer Vision and Image Understanding, 2008. doi:
10.1016/j.cviu.2007.09.012.

To accelerate the process of computing spectra of subgraphs the *spectral integral variation technique* is used in the same paper.

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

The use of the signless Laplacian eigenvalues looks even better.

Structural and spectral perturbations of graphs

Informally speaking, similar graphs can be understood as being obtained, one from the other, by some "small" perturbation. A graph perturbation can be described as a "small" change in its structure or in its spectrum. The spectral integral variation technique, described in the last section, is just an example involving *graph perturbations*. A graph perturbation means a small change in graph structure (e.g., adding an edge or a vertex). We are interested in changes in graph eigenvalues caused by a perturbation.

There is a chapter in the book

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

devoted to graph perturbations and corresponding changes in the spectrum.

The problem of protecting the privacy appears in *social networks* on the Internet (for example, Facebook) when studying general properties of an existing network. A way to protect the privacy of personal data is to randomize the network representing relations between individuals by deleting some actual edges and by adding some additional edges in such a way that the global characteristics of the network are unchanged.

This is achieved by using eigenvalues of the adjacency matrix (in particular, the largest one) and of the Laplacian (algebraic connectivity) to control the process of deleting and adding the edges, see
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.
The choice of deleted and added edges is performed by using results of *Eigenspaces of Graphs*, Chapter 6, for the largest eigenvalue and the corresponding results for the algebraic connectivity have been derived in the paper.

In Computer science literature, some spectral perturbations of graphs have been considered as well. This means that the graph spectrum is slightly changed while the eigenvectors remain unchanged. This is used in connection with the formula for spectral decomposition of the adjacency matrix A of a graph, i.e. $A = U\Lambda U^T$, where Λ is a diagonal matrix containing the eigenvalues of A and the columns of matrix U are orthonormal eigenvectors of A. The paper
Liu D., Wang H., Van Mieghem P., *Spectral perturbation and reconstructability of complex networks*, Phys. Rev. E, 81(2010), 016101, 1-9.
proposes a new robustness parameter for complex networks: this is the maximal number k such that one can replace k smallest in modulus eigenvalues of A with zeros with the possibility that A still can be reconstructed.

A similar "deletion" of eigenvalues appears also in the so called *latent semantic indexing* (LSI) but it is applied on singular values of the term-by-document matrix; see, e.g.,
Papadimitriou C.H., Raghavan P., Tamaki H., Vempala S., *Latent semantic indexing: a probabilistic analysis*, J. Comput. System Sci., 61(2000), No. 2, 217-235.

Some other spectral recognition problems

The second smallest Laplacian eigenvalue is called *algebraic connectivity* of the graph and was introduced by Fiedler

Fiedler M., *Algebraic connectivity of graphs*, Czechoslovak Math. J., 23(98)(1973), 298-305.

The eigenvector belonging to the second smallest Laplacian eigenvalue of the connected graph is called the *Fiedler eigenvector*. Of course, we have both positive and negative entries in it.

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

Fiedler M., *A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory*, Czechoslovak Math. J., 25(100)(1975), 619-633.

These ideas were exploited in the literature in various ways for devising powerful heuristics for spectral graph partitioning and/or clustering. For instance, Shi and Malik

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.

have shown how the sign pattern of the Fiedler eigenvector can be used to separate the foreground from the background structure in images. The original procedure Fiedler's has been improved by using the matrix $D^{-1}L$ (so as to maximize the normalized graph cut).

More generally, image segmentation is an important procedure in computer vision and pattern recognition. The problem is to divide the image into regions according to some criteria. Very frequently the image segmentation is obtained using eigenvectors of some graph matrices; for more details see, e.g.,
Weiss Y., *Segmentation using eigenvectors: a unifying view*, Proc. Seventh IEEE Internat. Conf. Computer Vision, 2(1999), pp. 975-982.

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

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

We consider the symmetric travelling salesman problem with instances I represented by complete graphs G with distances between vertices (cities) as edge weights (lengths). Intuitively, the hardness of an instance G depends on the distribution of short edges within G. Therefore we consider some short edge subgraphs of G (minimal spanning tree, critical connected subgraph, critical 2-connected subgraph and several others) as non-weighted graphs and several their invariants as potential complexity indices. Here spectral invariants (e.g. spectral radius of the adjacency matrix) play an important role since, in general, there are intimate relations between eigenvalues and the structure of a graph.

Since hidden details of short edge subgraphs really determine the hardness of the instance, one should use techniques of data mining to find them. In particular, spectral clustering algorithms are used including information obtained from the spectral gap in Laplacian spectra of short edge subgraphs:

Cvetković D., *Complexity indices for the travelling salesman problem and data mining*, Transactions of Combinatorics, 1(2012), No. 1, 35-43.

Thank you for your attention