

Matrix Representations of Graphs and Their Experimental Comparison for Detecting Non-subgraphs by Eigenvalues

KAORU KATAYAMA^{1,a)} TAKUMI SATO^{1,b)}

Received: March 23, 2014, Accepted: May 17, 2014

Abstract: Eigenvalues of graphs have been used for detecting non-subgraphs or non-supergraphs based on their interlacing property. However the detected subgraphs are often restricted to induced subgraphs or trees due to their matrix representations. We consider five matrix representations of a graph, which can be used to detect general non-subgraphs or non-supergraphs, and compare them experimentally.

Keywords: graph, eigenvalue, matrix

1. Introduction

Graphs are commonly used as models of real data such as XML documents, 3D objects, chemical compounds, Web data and social networks [1]. Eigenvalues of graphs are used in indexing [2], [3], [4] and clustering [5] of graphs. The interlacing property of eigenvalues [6] is applied to detecting non-subgraphs or non-supergraphs of a query graph from a graph database. Although the adjacency matrix is commonly used as a representation of a graph, that of a subgraph of the graph does not have the prerequisite for the property. That is, the adjacency matrix of the subgraph is not a principal submatrix of the adjacency matrix of the graph in general. In such cases, the detected subgraphs are restricted to induced subgraphs or trees due to their matrix representations. In order to detect general non-subgraphs, we define a matrix representation of a labeled graph in Ref. [7], which have the prerequisite of the theorem. However the size of the matrix is much larger than the size of the adjacency matrix and it costs more to compute the eigenvalues. In this paper, we present five matrix representations of a graph for detecting non-subgraphs or non-supergraphs, which include a matrix modified slightly from that used in Ref. [7]. They are based on the adjacency matrix, the signless Laplacian matrix and the Seidel matrix of a graph or its line graph. For non-regular graphs, there is no simple relation among the eigenvalues of these matrices [8]. We compare the processing time and the number of detected non-subgraphs experimentally for the matrix representations. To our knowledge, this is the first such experimental comparison of these matrix representation in terms of detecting non-subgraphs or non-supergraphs according to the interlacing property of the eigenvalues. In Ref. [7], it is shown that the processing time and the number of detected non-supergraphs are improved by decomposing graphs according

to labels of the vertices and the edges. We also decompose graphs as the same way before computing their eigenvalues in this work.

2. Preliminaries

2.1 Graphs and Matrices

We focus on a *labeled undirected simple graph* $g = (V^g, E^g, L(V)^g, L(E)^g, \mu^g, \nu^g)$. V^g is a set of vertices. E^g is a set of edges where an edge e in E^g is an ordered pair (v_1, v_2) of vertices v_1 and v_2 in V^g . $L(V)^g$ and $L(E)^g$ are sets of labels of vertices and edges, respectively. μ^g and ν^g are labeling functions $V^g \rightarrow L(V)^g$ and $E^g \rightarrow L(E)^g$, respectively. The *line graph* $\mathcal{L}(g)$ of a simple graph g is the graph where a vertex is associated with each edge of g and two vertices are adjacent if and only if the corresponding edges of g have a vertex in common. **Figure 1** shows examples of a graph and its line graph. We use the signless Laplacian matrix and the Seidel matrix of a graph in addition to the adjacency matrix and the incidence matrix as matrix representations of a graph. The *signless Laplacian matrix* P^g of a graph g is a $|V^g| \times |V^g|$ matrix where the i -th diagonal element is the degree of the vertex $v_i \in V^g$, the (i, j) element is 1 if $(v_i, v_j) \in E^g$ and the other elements are 0. The *Seidel matrix* $S^g = (s_{ij})$ of a graph g is a $|V^g| \times |V^g|$ matrix where the diagonal elements are 0, the (i, j) element is -1 if $(v_i, v_j) \in E^g$ and the other elements are 1.

2.2 Graphs and Eigenvalues

Let $\{\alpha_i\}_{i=1,\dots,n}$ and $\{\beta_j\}_{j=1,\dots,m}$ be two ordered sequences of real numbers where $m < n$, $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_n$ and $\beta_1 \leq \beta_2 \leq \dots \leq \beta_m$, respectively. We say that $\{\beta_j\}_{j=1,\dots,m}$ *interlaces* $\{\alpha_i\}_{i=1,\dots,n}$, if the condition, $\alpha_k \leq \beta_k \leq \alpha_{k+(n-m)}$, is satisfied for $k = 1, \dots, m$.

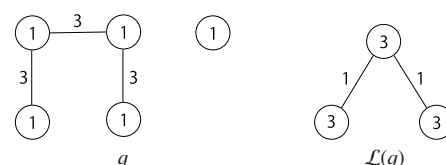


Fig. 1 Graph g and its line graph $\mathcal{L}(g)$.

¹ Graduate School of System Design, Tokyo Metropolitan University, Hino, Tokyo 191-0065, Japan

^{a)} kaoru@tmu.ac.jp

^{b)} sato-takumi@sd.tmu.ac.jp

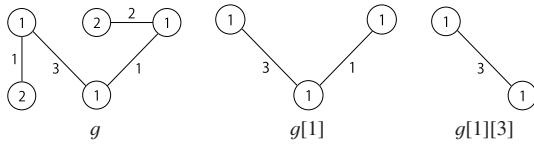


Fig. 2 Graph g and decomposed graphs $g[1], g[1][3]$.

Theorem 1 (Interlace Theorem) [6] Let A and B be $n \times n$ and $m \times m$ real symmetric matrices where $n > m$, respectively. If B is a principal submatrix of A , eigenvalues of B interlace those of A .

If matrix representations of a graph and its subgraph have the prerequisite of this theorem, we can check whether a graph is *not* a subgraph of another graph by comparing their eigenvalues.

2.3 Graph Decomposition by Labels

We decompose graphs according to labels of the vertices and the edges before computing their eigenvalues as follows, which is the same as Ref. [7].

VE decomposing graphs according to labels of vertices and then decomposing the result according to labels of edges

EV decomposing graphs according to labels of edges and then decomposing the result according to labels of vertices

Example 1 Figure 2 shows an example of the decomposition. A graph g is decomposed into $g[1]$ according to the label 1 of the vertex. $g[1]$ is decomposed into $g[1][3]$ according to the label 3 of the edge.

If g^s is a subgraph of g , the graphs into which g^s is decomposed according to a label are also subgraphs of the graphs into which g is decomposed according to the same label. In detecting non-subgraphs or non-supergraphs, eigenvalues of decomposed graphs which have the same labels are compared based on the theorem. Before computing the eigenvalues, we compare the sizes of a matrix representations of decomposed graphs of g and g^s . If that of g^s is larger than that of g with the same labels, g^s is not a subgraph of g . In such case, there is no need to compute their eigenvalues.

3. Matrix Representations for Detecting Non-subgraphs

We used the following matrix representation of a graph for detecting non-subgraphs based on the interlace theorem in Ref. [7]. The *extended incidence matrix* C^g of a graph g is the matrix $\begin{bmatrix} Q & N \\ {}^tN & R \end{bmatrix}$ where $N = (n_{ij})$ is the $|V^g| \times |E^g|$ incidence matrix of g , $Q = (q_{ij})$ and $R = (r_{ij})$ are $|V^g| \times |V^g|$ and $|E^g| \times |E^g|$ diagonal matrices whose i -th diagonal element is the label $\mu^g(v_i)$ of the vertex $v_i \in V^g$ and the label $\nu^g(e_i)$ of the edge $e_i \in E^g$, respectively. tN is the transpose of a matrix N . In this paper, we assign the label $\nu^g(e_j)$ of the edge $e_j \in E^g$ to the (i, j) element n_{ij} of N . In order to reduce the sizes of matrices and the cost of computing their eigenvalues, we consider the following smaller matrix representations.

Adjacency matrix of line graph Although the adjacency matrix of a subgraph g^s of a graph g is not the principal submatrix of the adjacency matrix of g in general, the adjacency matrix $A^{\mathcal{L}(g^s)}$ of the line graph $\mathcal{L}(g^s)$ is always the principal submatrix of the adjacency matrix $A^{\mathcal{L}(g)}$ of $\mathcal{L}(g)$. However g

cannot be restored from $\mathcal{L}(g)$ in general. We assign labels of vertices and edges of g to the corresponding edges and vertices of $\mathcal{L}(g)$ as the graphs of Fig. 1, respectively. The i -th diagonal element of $A^{\mathcal{L}(g)}$ is the label $\mu^{\mathcal{L}(g)}(v_i)$ of the vertex $v_i \in V^{\mathcal{L}(g)}$, the (i, j) element is the label $\nu^{\mathcal{L}(g)}(v_i, v_j)$ of the edge $(v_i, v_j) \in E^{\mathcal{L}(g)}$ and the other elements are 0.

Seidel matrix of line graph As with the adjacency matrix of a graph g , the Seidel matrix S^g of g does not have the prerequisite of the interlace theorem, but the Seidel matrix $S^{\mathcal{L}(g)}$ of the line graph $\mathcal{L}(g)$ does. Since there is no simple way to assign labels of vertices and edges of $\mathcal{L}(g)$ to elements of the Seidel matrix like the adjacency matrix, we do not assign the labels to the elements and use the Seidel matrix as the definition.

Signless Laplacian matrix The signless Laplacian matrix P^g of a graph g does not have the prerequisite of the interlace theorem. However we can use it as follows. P^g is equal to the multiplication $N \times {}^tN$ of the $|V^g| \times |E^g|$ incidence matrix N and its transpose tN . Although $P^g = N \times {}^tN$ does not have the prerequisite, the matrix ${}^tN \times N$ does. In addition, the non-zero eigenvalues of $N \times {}^tN$ are equal to those of ${}^tN \times N$. Therefore we can use eigenvalues of ${}^tN \times N$ or $P^g = N \times {}^tN$ to detecting non-subgraphs based on the theorem by adding a necessary number of zeros to them. In order to reduce the cost of computing eigenvalues, we choose the matrix whose size is smaller. We do not assign labels of the vertices and the edges to elements of ${}^tN \times N$ or $P^g = N \times {}^tN$ and use these matrices as the definition.

Signless Laplacian matrix of line graph If a graph g^s is a subgraph of a graph g , the line graph $\mathcal{L}(g^s)$ is also a subgraph of $\mathcal{L}(g)$. Therefore we can use the signless Laplacian matrix $P^{\mathcal{L}(g)}$ of the line graph in the same way as P^g .

Example 2 Figure 3 shows the extended incidence matrix C^g , the adjacency matrix $A^{\mathcal{L}(g)}$, the Seidel matrix $S^{\mathcal{L}(g)}$, the two signless Laplacian matrices $P^g = N \times {}^tN$, $P^{\mathcal{L}(g)}$, and the matrix ${}^tN \times N$ for the graph g and its line graph $\mathcal{L}(g)$ of Fig. 1. N is the incidence matrix of g . Since the size of ${}^tN \times N$ is smaller than $P^g = N \times {}^tN$ in this case, we compute the eigenvalues of ${}^tN \times N$ instead of those of $P^g = N \times {}^tN$. The eigenvalues of ${}^tN \times N$ and P^g are $\{2 - \sqrt{2}, 2, 2 + \sqrt{2}\}$ and $\{0, 0, 2 - \sqrt{2}, 2, 2 + \sqrt{2}\}$, respectively.

$$C^g = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 3 \\ 0 & 3 & 3 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 3 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 & 3 & 0 & 0 & 3 \end{bmatrix}, A^{\mathcal{L}(g)} = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix},$$

$$S^{\mathcal{L}(g)} = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix}, P^g = N \times {}^tN = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$${}^tN \times N = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}, P^{\mathcal{L}(g)} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Fig. 3 Matrix representations of graph g and its line graph $\mathcal{L}(g)$ of Fig. 1.

4. Experimental Evaluation

For the five matrix representations of a graph g and a smaller graph g^s , which are described in Section 3, we compare the processing time and the number of detected non-subgraphs in the cases of varying the number of vertices of g^s , the densities of g and g^s , and the number of labels of vertices of g and g^s . The processing time is the time required to compute eigenvalues of the decomposed graphs of g and g^s and compare them. We check whether g^s is not a subgraph of g using the following procedure.

- (1) decompose g and g^s according to labels of the vertices and the edges in each of the ways VE and EV, which are described in Section 2.3
- (2) compare the sizes of matrix representations of each of the decomposed graphs of g and g^s
- (3) check whether eigenvalues of each of the decomposed graphs of g^s interlace eigenvalues of the decomposed graph of g with the same label

The processing time and the required memory size depend on the sizes of matrix representations of g and g^s , and the number of labels of their vertices and edges since they are decomposed according to labels. In each experiment, we prepare 500 pairs of graphs g^s and g and repeat the experiment twice. These graphs are generated randomly with the graph generator which we develop. The results of the experiments are the average performance for two sets of 500 pairs of graphs. All algorithms are implemented using Matlab R2012b. The experiments done on a PC running Microsoft Windows 7 Professional with the Intel Core i3 3.3 GHz processor and 16 GB RAM. The ways VE and EV for decomposing graphs have little affect on the results in the case of using the same matrix representation.

Table 1 shows the result when we vary the number of vertices of g^s from 50 to 90. The number of vertices of g is 100. The den-

sity of each g and g^s is 0.4. We assign 3 labels to each of vertices and edges in g and g^s . As the number of vertices of g^s increases, the number of detected non-subgraphs increases in all matrix representations. This is because it becomes more difficult that eigenvalues of g^s interlace those of g as the difference between the sizes of their matrix representations decreases. The largest number of non-subgraphs is detected in the case of using adjacency matrices of their line graphs. However it differs only slightly from that in the case of using extended incidence matrices or signless Laplacian matrices. The least number of non-subgraphs is detected in the case of using Seidel matrices of their line graphs. We detect non-subgraphs the fastest in the case of using signless Laplacian matrices. This is due to the size of a signless Laplacian matrix of a graph, which depends on only the number of the vertices, is smaller than other matrices. The processing time for signless Laplacian matrices at 90 vertices becomes faster than that for 80 vertices since many non-subgraphs can be detected without computing their eigenvalues by comparing the sizes of each decomposed graphs in the second step of the procedure.

Table 2 shows the result when we vary the densities of g and g^s from 30 to 50. The numbers of vertices of g and g^s are 100 and 70, respectively. We assign 3 labels to each of vertices and edges of g and g^s . As the density of g^s decreases, the number of detected non-subgraphs increases in all the matrix representations. This is because difference between the sizes of matrix representations of g and g^s decreases as the density of g^s decreases. The number of detected non-subgraphs and the processing time for the five matrix representations shows the same tendency as the results shown in Table 1. The largest number of non-subgraphs is detected in the case of using adjacency matrices of their line graphs. As the densities of graphs increase, we can detect them much faster in the case of using signless Laplacian matrices than the case of using the other matrix representations.

Table 1 Varying number of vertices of smaller graph.

(a) Processing Time (sec)						(b) Number of Detected Non-subgraphs							
Vertices		50	60	70	80	90	Vertices		50	60	70	80	90
Extended Incidence Matrix	VE	6.34	6.79	7.61	8.78	10.05	VE	74	203	391	484	500	
	EV	6.44	6.88	7.68	8.79	9.97	EV	74	203	391	484	500	
Adjacency Matrix of Line Graph	VE	2.78	2.91	3.19	3.68	4.31	VE	79	204	391	484	500	
	EV	2.76	2.92	3.18	3.65	4.23	EV	79	204	391	484	500	
Seidel Matrix of Line Graph	VE	2.73	2.85	3.13	3.63	4.25	VE	56	159	348	472	500	
	EV	2.70	2.84	3.11	3.61	4.15	EV	56	159	348	472	500	
Signless Laplacian Matrix	VE	0.56	0.61	0.65	0.66	0.59	VE	75	203	391	484	500	
	EV	0.57	0.62	0.66	0.67	0.59	EV	75	203	391	484	500	
Signless Laplacian Matrix of Line Graph	VE	3.18	3.33	3.56	3.76	3.59	VE	72	195	383	478	500	
	EV	3.14	3.32	3.52	3.73	3.56	EV	72	195	383	478	500	

Table 2 Varying densities of pair of graphs.

(a) Processing Time (sec)						(b) Number of Detected Non-subgraphs							
Density		30	35	40	45	50	Density		30	35	40	45	50
Extended Incidence Matrix	VE	5.07	6.27	6.97	8.08	9.42	VE	427	385	376	355	340	
	EV	5.25	6.34	6.80	7.99	9.36	EV	427	385	376	355	340	
Adjacency Matrix of Line Graph	VE	1.73	2.39	2.92	3.68	4.60	VE	428	387	376	358	344	
	EV	1.72	2.35	2.81	3.59	4.53	EV	428	387	376	358	344	
Seidel Matrix of Line Graph	VE	1.69	2.34	2.85	3.61	4.52	VE	381	341	340	324	307	
	EV	1.69	2.30	2.76	3.52	4.42	EV	381	341	340	324	307	
Signless Laplacian Matrix	VE	0.62	0.67	0.66	0.68	0.71	VE	428	386	376	355	340	
	EV	0.64	0.69	0.65	0.68	0.71	EV	428	386	376	355	340	
Signless Laplacian Matrix of Line Graph	VE	1.91	2.74	3.42	4.37	5.56	VE	415	379	369	346	338	
	EV	1.90	2.70	3.32	4.30	5.54	EV	415	379	369	346	338	

Table 3 Varying number of vertex labels of pair of graphs.

(a) Processing Time (sec)							(b) Number of Detected Non-subgraphs						
Labels		2	3	4	5	6	Labels		2	3	4	5	6
Extended Incidence Matrix	VE	21.82	7.65	3.72	2.07	1.45	Extended Incidence Matrix	VE	132	384	478	498	500
	EV	21.51	7.59	3.84	2.13	1.60		EV	132	384	478	498	500
Adjacency Matrix of Line Graph	VE	11.86	3.20	1.32	0.73	0.53	Adjacency Matrix of Line Graph	VE	133	384	478	498	500
	EV	11.75	3.13	1.32	0.71	0.53		EV	133	384	478	498	500
Seidel Matrix of Line Graph	VE	11.67	3.14	1.29	0.71	0.51	Seidel Matrix of Line Graph	VE	119	336	458	495	500
	EV	11.57	3.08	1.28	0.69	0.51		EV	119	336	458	495	500
Signless Laplacian Matrix	VE	1.22	0.70	0.51	0.40	0.34	Signless Laplacian Matrix	VE	132	384	478	498	500
	EV	1.22	0.70	0.53	0.40	0.36		EV	132	384	478	498	500
Signless Laplacian Matrix of Line Graph	VE	14.00	3.73	1.42	0.69	0.45	Signless Laplacian Matrix of Line Graph	VE	145	380	470	496	500
	EV	13.99	3.64	1.41	0.68	0.44		EV	145	380	470	496	500

Table 3 shows the result when we vary the number of labels of vertices of g and g^s from 2 to 6. We assign 3 labels to edges of g and g^s . The numbers of vertices of g and g^s are 100 and 70, respectively. The densities of g and g^s are 0.4. As the number of labels of vertices increases, the number of detected non-subgraphs increases in all the matrix representations. This is because we check interlacing of eigenvalues of the more decomposed graphs of g and g^s as the the number of labels of vertices increases. In the case of using extended incidence matrices, adjacency matrices of their line graphs and signless Laplacian matrices, the largest number of non-subgraphs is detected for almost any number of labels of vertices. As the number of the labels decreases, we can detect them faster in the case of using signless Laplacian matrices than the case of using the other matrix representations. This is because the sizes of signless Laplacian matrices of decomposed graphs become smaller than the sizes of other matrix representations of them as the number of the labels decreases.

5. Conclusion

We present the five symmetric matrix representations of a graph g , which have the prerequisite of the interlace theorem. That is, the matrix representation of a subgraph g^s of g is a principal submatrix of the matrix representation of g , in general. We compare them experimentally in terms of the processing time and the number of non-subgraphs which are detected by comparing eigenvalues of g and g^s based on the theorem. The graphs are decomposed according to labels of the vertices and the edges before computing their eigenvalues. In the results of our experiment, the largest or almost the largest number of non-subgraphs are detected in the shortest time when we use the signless Laplacian matrix as the matrix representation.

References

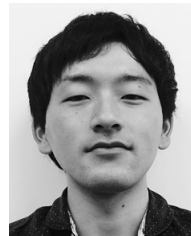
- [1] Cook, D.J. and Holder, L.B. (Eds.): *MINING GRAPH DATA*, John Wiley & Sons, Inc. (2007).
- [2] Zhang, N., Özsu, M.T., Ilyas, I.F. and Aboulnaga, A.: FIX: Feature-based Indexing Technique for XML Documents, *Proc. 32nd International Conference on Very Large Data Bases*, pp.259–270, VLDB Endowment (2006).
- [3] Shokoufandeh, A., Macrini, D., Dickinson, S., Siddiqi, K. and Zucker, S.W.: Indexing Hierarchical Structures Using Graph Spectra, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.27, No.7, pp.1125–1140 (2005).
- [4] Demirci, M.F., van Leuken, R.H. and Veltkamp, R.C.: Indexing through Laplacian spectra, *Computer Vision and Image Understanding*, Vol.110, No.3, pp.312–325 (2008).
- [5] Vinh, N., Inokuchi, A. and Washio, T.: Graph Classification Based on Optimizing Graph Spectra, *Discovery Science*, Lecture Notes in

Computer Science, Vol.6332, pp.205–220, Springer Berlin/Heidelberg (2010).

- [6] Haemers, W.H.: Interlacing Eigenvalues and Graphs, *Linear Algebra Appl.*, pp.593–616 (1995).
- [7] Katayama, K., Amagasa, Y. and Nagaya, H.: Detecting Non-subgraphs Efficiently by Comparing Eigenvalues of Decomposed Graphs, *IE-ICE Trans. Information and Systems*, Vol.E95-D, No.11, pp.2724–2727 (2012).
- [8] Cvetković, D., Rowlinson, P. and Simić, S.: *An Introduction to the Theory of Graph Spectra*, Cambridge University Press (2009).



Kaoru Katayama received his Ph.D. degree in informatics from Kyoto University in 2000. He is currently an associate professor in the Graduate School of Information and Communication Systems, Tokyo Metropolitan University. His research interests are in the areas of data engineering and data mining.



Takumi Sato received his Bachelor degree in engineering from Tokyo Metropolitan University in 2014. He is currently a master's student in the Graduate School of Information and Communication Systems, Tokyo Metropolitan University. His research interests are in the area of data engineering.