Int. J. Appl. Math. Comput. Sci., 2013, Vol. 23, No. 4, 823–837
DOI: 10.2478/amcs-2013-0062

A METHOD OF CONSTRUCTING THE FRAME OF A DIRECTED GRAPH

ICHIRO HOFUKU *, KUNIO OSHIMA **

* Laboratory of Mathematics
Tokyo Metropolitan College of Industrial Technology, Higashi-Ooi 140-0011, Japan
e-mail: hofuku@s.metro-cit.ac.jp

** School of Management
Tokyo University of Science, 500 Shimokiyoku 346-8512, Japan
e-mail: oshima@ms.kuki.sut.ac.jp

In web search engines, such as Google, the ranking of a particular keyword is determined by mathematical tools, e.g., Pagerank or Hits. However, as the size of the network increases, it becomes increasingly difficult to use keyword ranking to quickly find the information required by an individual user. One reason for this phenomenon is the interference of superfluous information with the link structure. The World Wide Web can be expressed as an enormous directed graph. The purpose of the present study is to provide tools for studying the web as a directed graph in order to find clues to the solution of the problem of interference from superfluous information, and to reform the directed graph to clarify the relationships between the nodes.

Keywords: directed graph, node clustering, Perron–Frobenius theorem, information retrieval.

1. Introduction

In web search engines, such as Google, the ranking of a particular keyword (called a query in the field of information retrieval (Berry *et al.*, 1999)) is determined by mathematical tools, e.g., Pagerank or Hits (Amy and Carl, 2005; 2008). However, as the size of the network increases, it becomes increasingly difficult to use keyword ranking to quickly find the information required by an individual user. One reason for this phenomenon is the interference of superfluous information with the link structure, i.e., meaningless links and redundant information, etc. In order to determine how to solve this problem, we note that the World Wide Web can be expressed as an enormous directed graph and then attack the following problems:

(1a) Cluster the nodes and generate a pair of groups of nodes for a given directed graph.

(1b) Provide a new method to simplify the structure of a given directed graph and construct the frame of such a graph using the diagrams generated in (1a).

Figure 1 shows examples where the relations between nodes are represented by directed edges.

The most common method of studying the simplification of a directed graph is to focus on the most strongly connected components or to focus on the distribution of the directed edges (Balakrishnan, 1997; Berge, 2001), and various methods of analysis have been performed (Aracena and Gomez, 2013; Yang *et al.*, 2012; Ligęza and Kościelny, 2008; Prelim and Demongeot, 2013). The method we propose (see (1a) and (1b)) is entirely different from the one that is currently being used. The structure of a directed graph is simplified by generating sets based on the degree of relation between a node n_i, which has a substantial significance, and a node n_j, which has substantial relations with n_i. Following the above processes (1a) and (1b), a given directed graph can be simplified and regarded as the frame of a directed graph for a given directed graph.

We use the following terminology (Amy and Carl, 2005). If node n_a has a directed edge pointing to node n_b, then we say that *node n_a is outlinked to node n_b* or that *node n_b is inlinked from node n_a* (see Fig. 2(a)). When node n_x is outlinked to other nodes, we say that *node n_x has outlinks*, and node n_x is referred to as the hub (see Fig. 2(b)). Conversely, when node n_y is inlinked

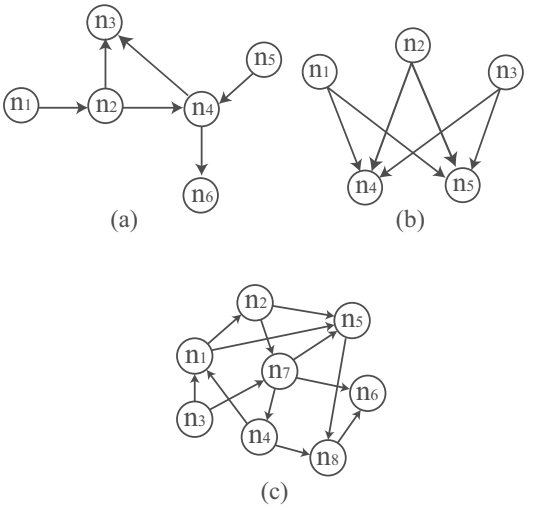


Fig. 1. Examples of a directed graph.

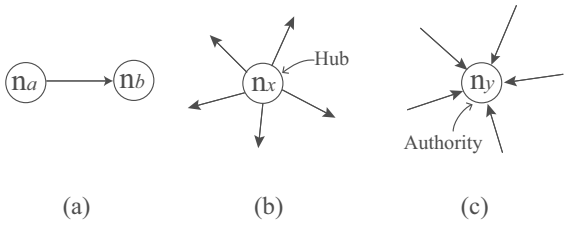


Fig. 2. Hub and authority.

from other nodes, we say that *node n_y has inlinks*, and node n_y is referred to as the authority (see Fig. 2(c)). A hub is a node that has many outlinks to authorities, and it corresponds to a web page such as link collection. An authority is a node that has "authority" over a given network, as the name suggests; and it contains a very large amount of information.

We also apply the following conditions.

Condition 1. *(Directed graph)* For our study, a directed graph must satisfy the following conditions:

(a) The directed graph is constructed from at least three nodes.

(b) All nodes of the directed graph are assumed to have at least one inlink or outlink.

Condition 1(a) means that, because of their simplicity, we will not consider directed graphs that have only two nodes. Condition 1(b) means that we will not consider directed graphs that contain independent nodes.

2. Previous study

In a previous study, we developed various models for ranking nodes (Hofuku and Oshima, 2012; 2010a; 2008; 2006). One of these, Ranking(I), is similar to the Pagerank and Hits algorithms. We combine the algorithms of Pagerank and Hits into a new algorithm, PH, in order to apply Ranking(I) to the web as a directed graph. We also consider a new ranking method, which is different from Pagerank and Hits, that determines the ranking based on the following two indices (Hofuku and Oshima, 2010b; Yokoi and Hofuku, 2010):

(2a) degree of significance between nodes,

(2b) degrees of relations between all the pairs of nodes.

While the rankings assigned by the Pagerank and Hits algorithms are based on the distribution of inlinks and outlinks among the nodes, the rankings assigned by the PH algorithm are based on degrees of relations between all the pairs of nodes that consider the distribution of inlinks along directed edges.

As mentioned above, the PH algorithm combines the Pagerank and Hits algorithms. We therefore consider the properties of these algorithms, as follows.

Properties of Pagerank. A page that is inlinked from several good pages is also a good page, and the rank of a page depends on the degree to which pages with several inlinks are linked to it. Figure 3(a) shows the properties of Pagerank; the page n_x which is inlinked from the page n_y, which has many inlinks, is a good page.

Properties of Hits. Unlike Pagerank, Hits has two kinds of nodes, authorities and hubs. Good hubs point to good authorities, and good authorities are pointed to by good hubs. A hub score and an authority score are assigned to each web page. Figure 3(b) shows the outline of the properties of Hits. Figure 3(b)(i) shows an authority, a page n_x, that is a good page because it is inlinked from the page n_y, which has many outlinks and is a good page. Figure 3(b)(ii) shows a hub, a page n_x, that is a good page because it is outlinked to the page n_y, which has many inlinks and is a good page.

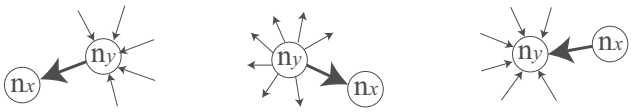
As mentioned above, Ranking(I) is needed to perform the PH algorithm. We present an overview of Ranking(I) in the next subsection.

2.1. Ranking(I).

In this subsection, a simple review of Ranking(I) is presented (for details, see Hofuku and Oshima, 2010a).

Let $\mathbf{M}_{(\mathrm{I})} = \{m_{(\mathrm{I})}[i,j]\}, 1 \le i,j \le n$ be a matrix generated by comparing two elements in

$$C = \{c(1), c(2), \ldots, c(n)\}$$



(i) The view of Authority (ii) The view of Hub

(a) Pagerank (b) Hits

Fig. 3. Properties of Pagerank and Hits.

through either competition or a trial. Each element in $\mathbf{M}_{(\mathrm{I})}$ is determined in accordance with the following conditions.

Condition 2.

(a) Matrix $\mathbf{M}_{(\mathrm{I})}$ is irreducible and primitive.

(b) The value of $m_{(\mathrm{I})}[i,j]$ represents the nonnegative ratio of superiority of $c(i)$ over $c(j)$.

(c) The ratio of superiority is determined by a common rule through either competition or a trial among the elements in C.

From Condition 2(b), no element of matrix $\mathbf{M}_{(\mathrm{I})}$ is negative, and so $\mathbf{M}_{(\mathrm{I})}$ is nonnegative. A matrix $\mathbf{M}_{(\mathrm{I})}$ that satisfies Condition 2 is referred to as an *evaluation matrix(I) corresponding to* C. Then, we have the following remark and definition.

Remark 1. From the Perron–Frobenius theorem (Berman and Plemmons, 1979; Lancaster and Tismenetsky, 1985) as well as Conditions 2(a) and 2(b), there exists an eigenvector $\boldsymbol{r}_{\mathrm{M}_{(\mathrm{I})}} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$, whose elements are all positive, which corresponds to the largest positive eigenvalue $\lambda_{\mathrm{M}_{(\mathrm{I})}}$ of $\mathbf{M}_{(\mathrm{I})}$.

Definition 1. The eigenvector $\boldsymbol{r}_{\mathrm{M}_{(\mathrm{I})}}$, given in Remark 1, is referred to as the *ranking vector corresponding to matrix* $\mathbf{M}_{(\mathrm{I})}$ and is normalized with respect to the l_2-norm.

In this study, each element in the initial vector is equal to 1 in the power method (Ortega, 1990), and $||\cdot||$ represents the l_2-norm. In the next section, the properties of each element in the ranking vector are given.

2.1.1. Process of generating the ranking vector for $\mathbf{M}_{(\mathrm{I})}$. In this subsection, we describe the process of generating the ranking vector and discuss the mathematical meaning of each element in it. From Condition 2(a), we can generate the ranking vector for $\mathbf{M}_{(\mathrm{I})}$ using the power method. Then, the initial vector is given as $\boldsymbol{r}_0 = (1, 1, \ldots, 1)^{\mathrm{T}}$, and

$$\mathbf{M}_{(\mathrm{I})}\boldsymbol{r}_0 \equiv \boldsymbol{r}_1 = \big(r_1(1),\ r_1(2),\ \ldots,\ r_1(n)\big)^{\mathrm{T}}. \tag{1}$$

From Eqn. (1), we define the *first potential vector* $\boldsymbol{p}_{[1]\mathrm{M}_{(\mathrm{I})}}$ for $\mathbf{M}_{(\mathrm{I})}$ as follows:

$$\begin{aligned}&\boldsymbol{p}_{[1]\mathrm{M}_{(\mathrm{I})}}\\ &= \frac{\boldsymbol{r}_1}{||\boldsymbol{r}_1||}\\ &= \big(p_{[1]\mathrm{M}_{(\mathrm{I})}}(1),\ p_{[1]\mathrm{M}_{(\mathrm{I})}}(2), \cdots,\ p_{[1]\mathrm{M}_{(\mathrm{I})}}(n)\big)^{\mathrm{T}}.\end{aligned} \tag{2}$$

An entry $p_{[1]\mathrm{M}_{(\mathrm{I})}}(i)$ in $\boldsymbol{p}_{[1]\mathrm{M}_{(\mathrm{I})}}$ is referred to as the *first potential for $c(i)$ in C*. Elements $p_{[1]\mathrm{M}_{(\mathrm{I})}}(i)$ $(i = 1, \ldots, n)$ in $\boldsymbol{p}_{[1]\mathrm{M}_{(\mathrm{I})}}$ represent the total degree of superiority of $c(i)$ to other elements (including the superiority of $c(i)$ to $c(i)$). Then, calculating $\mathbf{M}_{(\mathrm{I})}\boldsymbol{p}_{[1]\mathrm{M}_{(\mathrm{I})}}$, we obtain

$$\begin{aligned}&\mathbf{M}_{(\mathrm{I})}\boldsymbol{p}_{[1]\mathrm{M}_{(\mathrm{I})}}\\ &= \left(\sum_{k=1}^{n} m_{(\mathrm{I})}[1,k]\ p_{[1]\mathrm{M}_{(\mathrm{I})}}(k),\right.\\ &\qquad\left.\ldots, \sum_{k=1}^{n} m_{(\mathrm{I})}[n,k]\ p_{[1]\mathrm{M}_{(\mathrm{I})}}(k)\right)^{\mathrm{T}}\end{aligned} \tag{3}$$

$$\equiv \boldsymbol{r}_2 = \big(r_2(1),\ r_2(2),\ \ldots,\ r_2(n)\big)^{\mathrm{T}}. \tag{4}$$

Thus, if $c_2(i)(\in C)$ has a high rate of superiority to the other elements that have high first potentials, the value of $r_2(i)$ in $\boldsymbol{r}_2$ becomes characteristically larger than that of $r_2(j)$ $(i \neq j, 1 \leq j \leq n)$, where $c_2(j)$ has a high rate of superiority to the elements that have low first potentials compared with $c_2(i)$. As in the case of generating $\boldsymbol{p}_{[1]\mathrm{M}_{(\mathrm{I})}}$, $\boldsymbol{p}_{[2]\mathrm{M}_{(\mathrm{I})}}$ is also defined by normalizing $\boldsymbol{r}_2$, which is referred to as the *second potential vector*. Thus, the characteristic of $\{r_2(i)\}$ $(i = 1, \ldots, n)$ mentioned above is retained by $\{p_{[2]\mathrm{M}_{(\mathrm{I})}}(i)\}$, $(i = 1, \ldots, n)$. In a similar manner, $\boldsymbol{p}_{[3]\mathrm{M}_{(\mathrm{I})}}$, $\boldsymbol{p}_{[4]\mathrm{M}_{(\mathrm{I})}}$,... are defined as follows:

$$\begin{aligned}\mathbf{M}_{(\mathrm{I})}\boldsymbol{p}_{[\mathrm{k}-1]\mathrm{M}_{(\mathrm{I})}} &\equiv \boldsymbol{r}_k,\\ \boldsymbol{p}_{[\mathrm{k}]\mathrm{M}_{(\mathrm{I})}} &= \frac{\boldsymbol{r}_k}{||\boldsymbol{r}_k||},\end{aligned} \tag{5}$$

$k = 3, 4, \ldots$ and each entry in the k-th potential vector,

$$\boldsymbol{p}_{[\mathrm{k}]\mathrm{M}_{(\mathrm{I})}} = \big(p_{[\mathrm{k}]\mathrm{M}_{(\mathrm{I})}}(1),\ p_{[\mathrm{k}]\mathrm{M}_{(\mathrm{I})}}(2),\ \ldots,\ p_{[\mathrm{k}]\mathrm{M}_{(\mathrm{I})}}(n)\big)^{\mathrm{T}}, \tag{6}$$

has the following property.

Property 1. The k-th potential $p_{[\mathrm{k}]\mathrm{M}_{(\mathrm{I})}}(i)$ for $c(i)$, $1 \leq i \leq n$, which has a high rate of superiority to the other elements having high $(k-1)$-th potentials, becomes larger than the k-th potential $p_{[\mathrm{k}]\mathrm{M}_{(\mathrm{I})}}(j)$ for $c(j)$, $1 \leq j \leq n$, which have a high rate of superiority to the elements that have low $(k-1)$-th potentials.

The matrix $\mathbf{M}_{(\mathrm{I})}$ is assumed to be irreducible and primitive. From iterating as above, we can generate the ranking vector $\boldsymbol{r}_{\mathrm{M}_{(\mathrm{I})}}$, as defined in Definition 1,

corresponding to the largest positive eigenvalue $\lambda_{\mathrm{M}_{(\mathrm{I})}}$. This iteration process is identical to the generation of $\boldsymbol{r}_{\mathrm{M}_{(\mathrm{I})}}$ by the power method. Therefore, we have

$$\lim_{k\to\infty} \boldsymbol{p}_{[\mathrm{k}]\mathrm{M}_{(\mathrm{I})}} = \boldsymbol{r}_{\mathrm{M}_{(\mathrm{I})}}. \tag{7}$$

We refer to

$$\boldsymbol{p}_{[\infty]\mathrm{M}_{(\mathrm{I})}} \equiv \lim_{k\to\infty} \boldsymbol{p}_{[\mathrm{k}]\mathrm{M}_{(\mathrm{I})}} \tag{8}$$

as the *final potential* for $\mathbf{M}_{(\mathrm{I})}$. A vector $\boldsymbol{p}_{[\infty]\mathrm{M}_{(\mathrm{I})}}$ is generated through the successive transition of each step's potentials for all elements in C. Thus, we obtain another property for $\boldsymbol{r}_{\mathrm{M}_{(\mathrm{I})}}$ as follows.

Property 2. The value of $c(i)$ in $\boldsymbol{r}_{\mathrm{M}_{(\mathrm{I})}}$ is determined based on its superiority to elements that have relatively high potentials.

In the present paper, a ranking that is ordered according to the highest-value element in $\boldsymbol{r}_{\mathrm{M}_{(\mathrm{I})}}$ is referred to as Ranking(I) for $\mathbf{M}_{(\mathrm{I})}$ in C. We give an example of the application for Ranking(I) as follows.

Example 1. The superiority relation among $C = \{c(1), c(2), c(3)\}$ is given in the evaluation matrix $\mathbf{M}_{(\mathrm{I})1}$ as follows:

$$\mathbf{M}_{(\mathrm{I})1} = \begin{pmatrix} 9/10 & 3/10 & 9/10 \\ 8/10 & 5/10 & 7/10 \\ 7/10 & 5/10 & 8/10 \end{pmatrix}. \tag{9}$$

From simple calculus, the first potential vector $\boldsymbol{p}_{[1]\mathrm{M}_{(\mathrm{I})1}}$ is

$$\boldsymbol{p}_{[1]\mathrm{M}_{(\mathrm{I})1}} = (0.59612, 0.567733, 0.567733)^{\mathrm{T}}, \tag{10}$$

so the first potential of $c(1)$ is the highest, and the first potentials of $c(2)$ and $c(3)$ are the same. Next we calculate the second potential vector $\boldsymbol{p}_{[2]\mathrm{M}_{(\mathrm{I})1}}$ as follows:

$$\boldsymbol{p}_{[2]\mathrm{M}_{(\mathrm{I})1}} = (0.597128, 0.567898, 0.566506)^{\mathrm{T}}. \tag{11}$$

Among the entries in $\mathbf{M}_{(\mathrm{I})1}$, the values of superiority of $c(2)$ and $c(3)$ compared with $c(1)$ are different, and the first potential of $c(1)$ is the highest, as shown in Eqn. (10). In this case, the value of the second potential of $c(2)$ is higher than that of $c(3)$ in $\boldsymbol{p}_{[2]\mathrm{M}_{(\mathrm{I})1}}$ because the ratio of superiority of $c(2)$ is higher than that of $c(3)$ compared with $c(1)$. After this, the ranking does not change in the subsequent potential transitions $\boldsymbol{p}_{[3]\mathrm{M}_{(\mathrm{I})1}}$, $\boldsymbol{p}_{[4]\mathrm{M}_{(\mathrm{I})1}}$, ... Finally, we can generate the ranking vector as follows:

$$\boldsymbol{r}_{\mathrm{M}_{(\mathrm{I})1}} = (0.597102, 0.567967, 0.566465)^{\mathrm{T}}, \tag{12}$$

and from the elements in $\boldsymbol{r}_{\mathrm{M}_{(\mathrm{I})1}}$, Ranking(I) is

$$\text{First} \cdots c(1), \quad \text{Second} \cdots c(2), \quad \text{Third} \cdots c(3).$$

♦

3. PH algorithm

As mentioned in Section 2, a new ranking method can be developed by the PH algorithm, and the ranking by this method is determined based on degrees of relations between all the pairs of nodes that consider the distribution of inlinks along directed edges. Then, in the present paper, we improve the PH algorithm in order to be able to consider the distribution of outlinks (not only the distribution of inlinks) along directed edges. As a result, the PH algorithm is reconstructed that can generate indices for (2a) and (2b) considering the degree of relation between each pair of nodes from both the sides of the distribution of inlinks and outlinks along the directed edges among the nodes. So, the ranking by the PH algorithm has two aspects: one is derived from considering the distribution of inlinks, and we denote this by "for the authority"; the other is derived from considering the distribution of outlinks, and we denote this by "for the hub". Therefore, like the Hits algorithm, the PH algorithm has two scores. From these, we obtain a node-clustering algorithm and can thus solve the problems of (1a) and (1b).

In order to execute the Pagerank and Hits algorithms, it is necessary to create a matrix $\mathbf{N}$ in which the entries are determined by the relations between the nodes as follows

Definition 2. From the relations between the nodes in a directed graph, a matrix $\mathbf{N} = \{n[i,j]\}$, $1 \le i,j \le n$, is defined according to the following condition:

$$n[i,j] = \begin{cases} 1 & n_j \text{ is inlinked from } n_i, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

We present the PH algorithm as Algorithm 1. Its detailed review is presented in Fig. 4 along with the actual data. The value of k in PH 1 was assumed to be zero, and it was considered from the viewpoint of the authority in order to more easily display the flow of the PH algorithm. We explain the PH algorithm in two steps because PH 1 through PH 5 produces a new index that represents the ratio of relation between the nodes.

3.1. Explanation of PH 1 through PH 5. An explanation of PH 1 through PH 5 is presented below. From Fig. 4, matrices $\mathbf{M}(k=0)$ and $\mathbf{U}_{\mathrm{A}}$ can be generated as follows:

$$\mathbf{M} = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}, \tag{18}$$

Algorithm 1. PH.

PH 1: From a directed graph, a matrix $\mathbf{M} = \{m[i,j]\}, 1 \le i,j \le n$, representing the relations of the nodes is defined as

$$\mathbf{M} = \mathbf{N} + k\,\mathbf{N}^2, \tag{14}$$

where k is a real parameter satisfying $0 \le k \le 1$ that controls the degree of influence of those directed edges that point from n_i to n_j at a distance of no more than two steps.

PH 2: Two matrices $\mathbf{U_A}$ and $\mathbf{U_H}$ are defined respectively as follows:

$$\mathbf{U_A} = {}^{\mathrm{T}}\mathbf{M}\,\mathbf{M}, \qquad \mathbf{U_H} = \mathbf{M}\,{}^{\mathrm{T}}\mathbf{M}, \tag{15}$$

where $\mathbf{U_A}$ corresponds to an authority and $\mathbf{U_H}$ corresponds to a hub.

PH 3: Two matrices, $\mathbf{V}_{\mathrm{A}} = \{v_{\mathrm{A}}[i,j]\}$ and $\mathbf{V}_{\mathrm{H}} = \{v_{\mathrm{H}}[i,j]\}$, are constructed by normalizing the rows of $\mathbf{U}_{\mathrm{A}}$ and $\mathbf{U}_{\mathrm{H}}$, respectively, with respect to the l_1-norm.

PH 4: When all the entries in a row of $\mathbf{V}_{\mathrm{A}}$ or $\mathbf{V}_{\mathrm{H}}$ are zero, we add a constant value to all of the entries in the row so that the sum of the row is equal to 1. The matrices $\mathbf{V}_{\mathrm{A}_1} = \{v_{\mathrm{A}_1}[i,j]\}$ and $\mathbf{V}_{\mathrm{H}_1} = \{v_{\mathrm{H}_1}[i,j]\}$ are both obtained through these procedures.

PH 5: In order to guarantee the irreducibility of $\mathbf{V}_{\mathrm{A}_1}$ and $\mathbf{V}_{\mathrm{H}_1}$, a tuning number c, $0 < c < 1$, modifies matrices $\mathbf{V}_{\mathrm{A}_1}$ and $\mathbf{V}_{\mathrm{H}_1}$ as follows:

$$\begin{aligned} \mathbf{V}'_{\mathrm{A}_1} &= (1-c)\frac{1}{n}\mathbf{E} + c\mathbf{V}_{\mathrm{A}_1}, \\ \mathbf{V}'_{\mathrm{H}_1} &= (1-c)\frac{1}{n}\mathbf{E} + c\mathbf{V}_{\mathrm{H}_1}, \end{aligned} \tag{16}$$

where $\mathbf{E}$ is the $n \times n$ matrix whose elements are all equal to 1.

PH 6: Set

$$\mathbf{W}_{\mathrm{A}} = {}^{\mathrm{T}}\mathbf{V}'_{\mathrm{A}_1}, \quad \mathbf{W}_{\mathrm{H}} = {}^{\mathrm{T}}\mathbf{V}'_{\mathrm{H}_1}. \tag{17}$$

PH 7: Generate the eigenvectors $\boldsymbol{r}_{\mathrm{W_A}}$ and $\boldsymbol{r}_{\mathrm{W_H}}$ for $\mathbf{W}_{\mathrm{A}}$ and $\mathbf{W}_{\mathrm{H}}$, respectively.

$$\mathbf{U}_{\mathrm{A}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 3 & 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{19}$$

As an example, consider the third row of matrix $\mathbf{U}_{\mathrm{A}} = \{u_{\mathrm{A}}[i,j]\}$. Entry $u_{\mathrm{A}}[3,j]$ of the third row represents the number of outlinks to n_j from nodes having outlinks to n_3 at a distance of one step where the value of k is assumed to be 0. The nodes having outlinks to n_3 are $\{n_1, n_2, n_6, n_7\}$, and the total number of outlinks is eight. Four of these nodes are toward n_3, one of these nodes is toward n_4, and three of these nodes are toward n_5. Thus, the results for each value are $u_{\mathrm{A}}[3,3] = 4$, $u_{\mathrm{A}}[3,4] = 1$, and finally, $u_{\mathrm{A}}[3,5] = 3$.

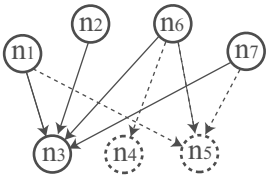


Fig. 4. Example of a directed graph.

For the matrix $\mathbf{U}_{\mathrm{A}}$, a matrix $\mathbf{V}_{\mathrm{A}} = \{v_{\mathrm{A}}[i,j]\}$ is created in PH3 as follows:

$$\mathbf{V_A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4/8 & 1/8 & 3/8 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 3/7 & 1/7 & 3/7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{20}$$

Note that in $\mathbf{V}_{\mathrm{A}}$ each column corresponds to a single node. As an example, if the third column corresponding to node n_3 is considered, the value $v_{\mathrm{A}}[5,3] = 3/7$ represents the ratio of outlinks to n_3 to the total number of nodes having outlinks to n_5. This value is an index of the quantity of relations between n_3 and n_5. This value becomes large as the relations between nodes n_3 and n_5 increase. This is natural because if the outlinks to nodes n_3 and n_5 (for the authority) are similar, then the quantity of $v_{\mathrm{A}}[5,3]$ between n_3 and n_5 increases. Therefore, the entries of the third column indicate the ratios of relation between node n_3 and the other nodes. We thus have the following property of $\mathbf{V}_{\mathrm{A}}$.

Property 3. Let the i-th column vector in $\mathbf{V}_{\mathrm{A}}$ be $\boldsymbol{v}_{\mathrm{A}}(i)$, $1 \le i \le n$. Then, each entry in $\boldsymbol{v}_{\mathrm{A}}(i)$ represents the ratio of relation between n_i and other nodes for the case in which $k = 0$ in PH 1.

Since matrix $\mathbf{V}'_{\mathrm{A}_1}$ in the processes of PH 4 and PH 5 is created only in order to guarantee the irreducibility of $\mathbf{V}_{\mathrm{A}}$, Property 3 of $\mathbf{V}_{\mathrm{A}}$ is inherited by that of $\mathbf{V}'_{\mathrm{A}_1}$.

3.2. Degree of relation between nodes. For PH 1 through PH 5, it is possible to define new indices that represent the degree of relation between any pair of nodes for the authority and the hub for a given k in PH 1.

Definition 3. For the i-th and j-th columns in $\mathbf{V}'_{\mathrm{A}_1}$, denoted by $\boldsymbol{v}'_{\mathrm{A}_1}(i)$ and $\boldsymbol{v}'_{\mathrm{A}_1}(j)$, respectively, $1 \le i, j \le n$, the following index and matrix are defined:

(a) *(Ratio of relation between two nodes for the authority)* The ratio of relation between node n_i and n_j for the authority is defined as follows:

$$r_{\mathrm{A}}(i,j;k) = \frac{\boldsymbol{v}'_{\mathrm{A}_1}(i) \cdot \boldsymbol{v}'_{\mathrm{A}_1}(j)}{||\boldsymbol{v}'_{\mathrm{A}_1}(i)||\ ||\boldsymbol{v}'_{\mathrm{A}_1}(j)||}, \qquad (21)$$

where $\cdot$ denotes the inner product.

(b) *(Node-relation matrix for the hub)* The matrix $\mathbf{R}_{(\mathrm{A};k)}$, with entries $\{r_{\mathrm{A}}(i,j;k)\}$, is referred to as the node-relation matrix for the authority.

In Definition 3(a), the value $r_{\mathrm{A}}(i,j;k)$ in Eqn. (21) is $\cos\theta$, where θ is the angle between vectors $\boldsymbol{v}'_{\mathrm{A}_1}(i)$ and $\boldsymbol{v}'_{\mathrm{A}_1}(j)$, and $\boldsymbol{v}'_{\mathrm{A}_1}(i)$ represents the ratios of relation between n_i and the other nodes for the authority. Therefore, $r_{\mathrm{A}}(i,j;k)$ in Definition 3(a) represents an index that is determined by the similarity of the ratio of distribution in terms of inlinks to n_i and n_j from all nodes. Thus, the following property is obtained.

Property 4.

(a) The value $r_{\mathrm{A}}(i,j;k)$ is $0 \le r_{\mathrm{A}}(i,j;k) \le 1$, and the ratio of relation between n_i and n_j increases as $r_{\mathrm{A}}(i,j;k)$ approaches 1.

(b) The value of $r_{\mathrm{A}}(i,j;k)$ increases as the value of k in PH 1 increases.

In a manner similar to the process for the authority, the ratio of relation between n_i and n_j with respect to the hub can be defined as follows.

Definition 4. For the i-th and j-th columns in $\mathbf{V}'_{\mathrm{H}_1}$, denoted by $\boldsymbol{v}'_{\mathrm{H}_1}(i)$ and $\boldsymbol{v}'_{\mathrm{H}_1}(j)$, respectively, $1 \le i, j \le n$, the following index and matrix are defined:

(a) *(Ratio of relation between two nodes for the hub)* The ratio of relation between node n_i and n_j for the hub is defined as follows:

$$r_{\mathrm{H}}(i,j;k) = \frac{\boldsymbol{v}'_{\mathrm{H}_1}(i) \cdot \boldsymbol{v}'_{\mathrm{H}_1}(j)}{||\boldsymbol{v}'_{\mathrm{H}_1}(i)||\ ||\boldsymbol{v}'_{\mathrm{H}_1}(j)||}. \qquad (22)$$

(b) *(Node-relation matrix for the hub)* The matrix $\mathbf{R}_{(\mathrm{H};k)}$, with entries $\{r_{\mathrm{H}}(i,j;k)\}$, is referred to as the node-relation matrix for the hub.

Property 5.

(a) The value $r_{\mathrm{H}}(i,j;k)$ is $0 \le r_{\mathrm{H}}(i,j;k) \le 1$, and the ratio of relation between n_i and n_j increases as $r_{\mathrm{H}}(i,j;k)$ approaches 1.

(b) The value of $r_{\mathrm{H}}(i,j;k)$ increases as the value of k in PH 1 increases.

3.2.1. Review of PH 6 and PH 7. We now discuss the mathematical meaning of PH 6 and PH 7. Entries $[i,j]$ in $\mathbf{W}_{\mathrm{A}}$ and $\mathbf{W}_{\mathrm{H}}$ in PH 6, $1 \le i, j \le n$, represent the ratio of the relation between n_i and n_j with respect to the authority and the hub, respectively. Thus, in the process of generating the eigenvectors of $\mathbf{W}_{\mathrm{A}}$ and $\mathbf{W}_{\mathrm{H}}$, each entry of $\boldsymbol{r}_1$ in Eqn. (1) is the ratio of the relation between itself and the other nodes. Therefore, each entry in the generated eigenvectors (ranking vector) $\boldsymbol{r}_{\mathrm{W_A}}$ and $\boldsymbol{r}_{\mathrm{W_H}}$, corresponding to the simple spectral radius, has the following property.

Property 6. The values of each entry in the ranking vectors $\boldsymbol{r}_{\mathrm{W_A}}$ and $\boldsymbol{r}_{\mathrm{W_H}}$ depend on the ratios of relation to other nodes that, in turn, have relatively high ratios of relation to other nodes.

3.3. Application of the PH algorithm. In this subsection, we present the results of applying the PH algorithm to the graph in Fig. 1(a). Table 1 shows the resultant rankings for the PH algorithm with respect to the authority and the hub. Equations (23) and (24) present the node-relation matrices with respect to the authority in the cases $k = 0$ and $k = 0.5$, respectively:

$$\mathbf{R}_{(\mathrm{A};0)} = \begin{pmatrix} 1. & 0.318 & 0.419 & 0.421 & 1. & 0.495 \\ 0.318 & 1. & 0.126 & 0.128 & 0.318 & 0.152 \\ 0.419 & 0.126 & 1. & 0.687 & 0.419 & 0.883 \\ 0.421 & 0.128 & 0.687 & 1. & 0.421 & 0.326 \\ 1. & 0.318 & 0.419 & 0.421 & 1. & 0.495 \\ 0.495 & 0.152 & 0.883 & 0.326 & 0.495 & 1. \end{pmatrix}, \qquad (23)$$

$$\mathbf{R}_{(\mathrm{A};0.5)} = \begin{pmatrix} 1. & 0.520 & 0.436 & 0.520 & 1. & 0.585 \\ 0.520 & 1. & 0.578 & 0.700 & 0.520 & 0.384 \\ 0.436 & 0.578 & 1. & 0.961 & 0.436 & 0.929 \\ 0.520 & 0.700 & 0.961 & 1. & 0.520 & 0.852 \\ 1. & 0.520 & 0.436 & 0.520 & 1. & 0.585 \\ 0.585 & 0.384 & 0.929 & 0.852 & 0.585 & 1. \end{pmatrix}. \qquad (24)$$

Table 1. Rankings using the PH algorithm for Fig. 1(a).

	PH algorithm ($k = 0.5$)			
node	Value($\boldsymbol{r}_{\mathrm{A}}$)	Rank(Aut.)	Value($\boldsymbol{r}_{\mathrm{H}}$)	Rank(Hub)
n_1	0.0465	5	0.348	4
n_2	0.225	4	0.693	1
n_3	0.717	1	0.0482	5
n_4	0.524	2	0.453	2
n_5	0.0465	5	0.435	3
n_6	0.396	3	0.0482	5

4. Node clustering

In this section, we present a method for clustering the nodes of a given directed graph. First, using the indices of significance and the relations between the nodes based on the PH algorithm, we generate the authority set $\mathrm{D_A}$, the hub set $\mathrm{D_H}$, and the relay set $\mathrm{D_R}$ (as discussed below). In the following subsection, we present a method that uses a probability law to generate $\mathrm{D_A}$ and $\mathrm{D_H}$.

4.1. Introducing probability. First, we consider a trial that selects one node as first among all of the nodes in a directed graph. Let $\mathrm{P}(A_i),\ (i = 1,\dots,n)$, be the distribution probability for A_i, where node n_i is selected first. Next, let $\mathrm{P}(B_j),\ (j = 1,\dots,n)$, be the distribution probability for B_j, where node n_j is chosen second. In this case, node n_j is assumed to be chosen by sampling with replacement. Then, for n_i and n_j, the conditional probability is defined as follows:

$$\mathrm{P}(B_j|A_i) = \frac{\mathrm{P}(A_i, B_j)}{\mathrm{P}(A_i)}. \tag{25}$$

In the following subsection, we discuss the use of $\mathrm{P}(B_j|A_i)$.

4.2. Probability for generating the authority set. In this subsection, we present a method to generate the authority set, using the laws of conditional probability, from the matrix $\mathbf{R}_{(\mathrm{A};k)}$ and ranking vector $\boldsymbol{r}_{\mathrm{W_A}}$ that were generated by the PH algorithm with respect to the authority. The entry $r_{\mathrm{A}}(i,j;k)$ in matrix $\mathbf{R}_{(\mathrm{A};k)}$ represents the ratio of relation between n_i and n_j only. Then, focusing on node n_i, we may ask how we can define the ratios of relation for node n_i to each of the $\{n_j\}$. If the ratios of relation between n_i and each of the $\{n_j\}$ are large, the ratio of relation for the focused n_i to each of the $\{n_j\}$ will appear relatively small. Then, the index that represents the ratio of relation for the n_i of interest for each n_j $(j = 1, 2, \dots, n)$, called the connection of n_i to n_j, is defined as the conditional probability as follows:

$$\mathrm{P_A}(B_j|A_i) = \frac{1}{\sum_{j=1}^{n} r_{\mathrm{A}}(i,j;k)}\, r_{\mathrm{A}}(i,j;k), \qquad (1 \le i,j \le n). \tag{26}$$

For $\mathrm{P}(A_i)$ in Eqn. (25), using the ranking vector

$$\boldsymbol{r}_{\mathrm{W_A}} = \big(r_{\mathrm{W_A}}(1),\dots,r_{\mathrm{W_A}}(i),\dots,r_{\mathrm{W_A}}(n)\big)^{\mathrm{T}}$$

in $\mathbf{W}_{\mathrm{A}}$, we have the following equation:

$$\mathrm{P_A}(A_i) = \frac{1}{\sum_{i=1}^{n} r_{\mathrm{W_A}}(i)}\, r_{\mathrm{W_A}}(i), \quad (1 \le i \le n). \tag{27}$$

From Eqns. (26) and (27), the following equation is satisfied:

$$\mathrm{P_A}(A_i, B_j) = \mathrm{P_A}(A_i)\,\mathrm{P_A}(B_j|A_i). \tag{28}$$

In Eqn. (28), $\mathrm{P_A}(A_i, B_j)$ represents the joint distribution of $\{A_i\}$ and $\{B_j\}$.

The indices (2a) and (2b) in Section 2 are critical for grasping the structure of a directed graph. The value of $\mathrm{P_A}(A_i)$ increases if the node n_i is significant, and $\mathrm{P_A}(B_j|A_i)$ is the ratio of the relation for the focused n_i to n_j. Then the value of $\mathrm{P_A}(A_i, B_j)$ increases if the values of $\mathrm{P_A}(A_i)$ and $\mathrm{P_A}(B_j|A_i)$ both increase. The relation between an n_i that has substantial significance and n_j that has a substantial relation with n_i appears to substantially influence the structure of a given directed graph. Therefore, the value of $\mathrm{P_A}(A_i, B_j)$ in Eqn. (28) is referred to as the ratio of influence of the relation from n_i to n_j on a structure with respect to the authority. We denote the relation from n_i to n_j described above as $n_i \circ\!\!\rightarrow n_j$.

4.3. Probability for generating the hub set. In this subsection, using a law of conditional probability, we present a method for generating the hub set from $\mathbf{R}_{(\mathrm{H};k)}$ and a ranking vector $\boldsymbol{r}_{\mathrm{W_H}}$, which were generated from the PH algorithm with respect to the hub. This method is performed in the same manner in which we generated the authority set in Section 4.2. For a particular A_i and B_j, $\mathrm{P_H}(B_j|A_i)$ and $\mathrm{P_H}(A_i)$ are defined as follows:

$$\mathrm{P_H}(B_j|A_i) = \frac{1}{\sum_{j=1}^{n} r_{\mathrm{H}}(i,j;k)}\, r_{\mathrm{H}}(i,j;k), \tag{29}$$

$$\mathrm{P_H}(A_i) = \frac{1}{\sum_{i=1}^{n} r_{\mathrm{W_H}}(i)}\, r_{\mathrm{W_H}}(i), \qquad (1 \le i,j \le n). \tag{30}$$

From Eqns. (29) and (30), the joint distribution for $\{A_i\}$ and $\{B_j\}$ is defined as follows:

$$\mathrm{P_H}(A_i, B_j) = \mathrm{P_H}(A_i)\,\mathrm{P_H}(B_j|A_i). \tag{31}$$

As in the case of the authority, the value of $\mathrm{P_H}(A_i, B_j)$ in Eqn. (31) is referred to as the ratio of influence of the relation from n_i to n_j on a structure in terms of the hub (the situation of considering the outlinks among the nodes). We denote the relation from n_i to n_j described above as $n_i \bullet\!\!\rightarrow n_j$.

4.4. Method of generating the authority set and the hub set. We now present a method for generating the authority set $\mathrm{D_A}$ and the hub set $\mathrm{D_H}$, based on the values of $\mathrm{P_A}(A_i, B_j)$ and $\mathrm{P_H}(A_i, B_j)$ that were given in Sections 4.2 and 4.3. First, two stochastic matrices $\mathbf{T}_{\mathrm{A}}$ and $\mathbf{T}_{\mathrm{H}}$ are prepared from $\mathrm{P_A}(A_i, B_j)$ and $\mathrm{P_H}(A_i, B_j)$ as follows:

$$\begin{aligned} \mathbf{T}_{\mathrm{A}} &= \{t_{\mathrm{A}}[i,j]\} = \{\mathrm{P_A}(A_i, B_j)\}, \\ \mathbf{T}_{\mathrm{H}} &= \{t_{\mathrm{H}}[i,j]\} = \{\mathrm{P_H}(A_i, B_j)\}, \\ &\qquad (1 \le i,j \le n). \end{aligned} \tag{32}$$

Based on the characteristics of the process for generating $\mathbf{T}_{\mathrm{A}}$ and $\mathbf{T}_{\mathrm{H}}$ with Eqn. (32), $\mathbf{T}_{\mathrm{A}}$ and $\mathbf{T}_{\mathrm{H}}$ have the following properties:

$$\sum_{i,j=1}^{n} t_{\mathrm{A}}[i,j] = \sum_{i,j=1}^{n} t_{\mathrm{H}}[i,j] = 1,$$
$$\max_{j=1}^{n} t_{\mathrm{A}}[i,j] = t_{\mathrm{A}}[i,i], \quad \max_{j=1}^{n} t_{\mathrm{H}}[i,j] = t_{\mathrm{A}}[i,i], \quad (1 \le i,j \le n). \tag{33}$$

Then, based on each value of $\{t_{\mathrm{A}}[i,j]\}$ and $\{t_{\mathrm{H}}[i,j]\}$, the authority set $\mathrm{D_A}$, the hub set $\mathrm{D_H}$, and the relay set $\mathrm{D_H}$ are generated from the clustering algorithm as follows.

Algorithm 2. Node-clustering.

CL 1: *(Initial step of clustering)* The element with the largest order is selected from among $\{t_{\mathrm{A}}[i,j]\}$ and $\{t_{\mathrm{H}}[i,j]\}$, and the relation between the two nodes corresponding to the element is based on four patterns as follows:

$$\begin{cases} \text{(a) element } \textit{belonging to } \{t_{\mathrm{A}}[i,j]\}_{i\neq j} \\ \qquad\qquad \Longrightarrow\ n_i \circ\!\!\rightarrow n_j; \\ \text{(b) element } \textit{belonging to } \{t_{\mathrm{A}}[i,i]\} \\ \qquad\qquad \Longrightarrow\ n_i \circ\ ; \\ \text{(c) element } \textit{belonging to } \{t_{\mathrm{H}}[i,j]\}_{i\neq j} \\ \qquad\qquad \Longrightarrow\ n_i \bullet\!\!\rightarrow n_j; \\ \text{(d) element } \textit{belonging to } \{t_{\mathrm{H}}[j,j]\} \\ \qquad\qquad \Longrightarrow\ n_j \bullet\ . \end{cases}$$

If a selected element corresponds to (a) in CL 1, then the notation $n_i \circ\!\!\rightarrow n_j$ indicates that a directed edge has been created between n_i and n_j, as in Fig. 5(a). In a similar manner, (b) through (d) indicate situations as shown in Figs. 5(b), (c), and (d), respectively. If two or more relations between nodes exist for a certain value among the $\{t_{\mathrm{A}}[i,j]\}$ and $\{t_{\mathrm{H}}[i,j]\}$, the nodes corresponding to each element are denoted simultaneously. We now define an authority set and a hub set as follows:

Definition 5. (*Authority set and hub set*)

(a) If the relations $n_i \circ\!\!\rightarrow n_j$ and $n_j \circ\!\!\rightarrow n_i$ between two nodes n_i and n_j are satisfied, then we generate a new set $\mathrm{D_A} = \{n_i, n_j\}$, which is referred to as an *authority set*.

(b) If the relations $n_i \bullet\!\!\rightarrow n_j$ and $n_j \bullet\!\!\rightarrow n_i$ between two nodes n_i and n_j are satisfied, then we generate a new set $\mathrm{D_H} = \{n_i, n_j\}$, which is referred to as a *hub set* (see Fig. 6(b)).

In performing CL 1, the nodes that belong to an authority set or a the hub set are updated sequentially,

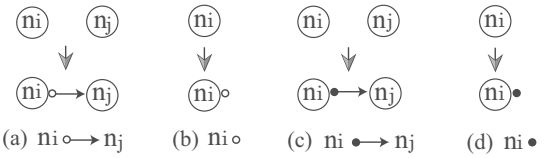


Fig. 5. Patterns from (a) through (d) in CL 1.

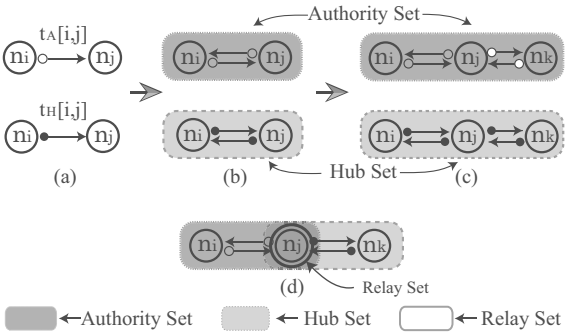


Fig. 6. Authority and hub sets.

as shown in Fig. 6(c). We thus denote the first generated authority set and hub set in CL 1 as $\mathrm{D}_{\mathrm{A}}^{(1)}$ and $\mathrm{D}_{\mathrm{H}}^{(1)}$, respectively, and if there exist two or more authority sets or hub sets, we denote each set as $\mathrm{D}_{\mathrm{A}}^{(2)}$, $\mathrm{D}_{\mathrm{A}}^{(3)}$, ... or $\mathrm{D}_{\mathrm{H}}^{(2)}$, $\mathrm{D}_{\mathrm{H}}^{(3)}$,$\cdots$, respectively, in the order created or in the order updated.

We now also define a relay node as follows.

Definition 6. (*Relay set*) A node $n_r \in \mathrm{D_A} \cap \mathrm{D_H}$ is referred to as a relay node, and a set of relay nodes, denoted by $\mathrm{D_R} = \mathrm{D_A} \cap \mathrm{D_H}$, is referred to as a relay set (see Fig. 6(d)).

CL 2: If, in CL 1, denoting the relations between nodes belonging to $\{t_{\mathrm{A}}[i,j]\}$ or $\{t_{\mathrm{H}}[i,j]\}$, there is a node that satisfies (a) or (b) as given below, then stop.

(a) There exist $n_x \in \mathrm{D_A}$ and $n_y \in \mathrm{D_H} - \mathrm{D_R}$ that satisfy the following condition: $n_x \circ\!\!\rightarrow n_y$.

(b) There exist $n_x \in \mathrm{D_H}$ and $n_y \in \mathrm{D_A} - \mathrm{D_R}$ that satisfy the following condition: $n_x \bullet\!\!\rightarrow n_y$.

In CL 2, the existence of n_x and n_y allows us to judge whether a boundary between n_x and n_y has been exceeded. The authority set and the hub set are characteristically different types of sets with respect to the generating process. Thus, (a) and (b) in CL 2 refer to states in which a superfluous relation between n_x and n_y arises.

CL 3: *(First step of clustering)* By performing CL 1 and CL 2, the initial steps of clustering are completed. We then cluster the remaining nodes, which were not clustered

in the initial step, in the same manner (using CL 1 and CL 2). In performing CL 3, if a new authority set or hub set is generated in this step, the new set is denoted by $\mathrm{D}_{\mathrm{A}[1]}$ or $\mathrm{D}_{\mathrm{H}[1]}$ to distinguish it from the authority set or the hub generated in the initial step. As in CL 1, if a node n_z exists that satisfies $n_z \in \mathrm{D}_{\mathrm{A}[1]} \cap \mathrm{D}_{\mathrm{H}[1]}$, then we define the set $\mathrm{D}_{\mathrm{R}[1]} = \mathrm{D}_{\mathrm{A}[1]} \cap \mathrm{D}_{\mathrm{H}[1]}$, $n_z \in \mathrm{D}_{\mathrm{R}[1]}$. In performing CL 3, if the nodes that belong to an authority set or a the hub set are updated sequentially, we denote the first generated authority set and hub set in CL 3 as $\mathrm{D}^{(1)}_{\mathrm{A}[1]}$ and $\mathrm{D}^{(1)}_{\mathrm{H}[1]}$, respectively. If there exist two or more authority sets or hub sets in the first step, we denote each set as $\mathrm{D}^{(2)}_{\mathrm{A}[1]}, \mathrm{D}^{(3)}_{\mathrm{A}[1]}, \ldots$ or $\mathrm{D}^{(2)}_{\mathrm{H}[1]}, \mathrm{D}^{(3)}_{\mathrm{H}[1]}, \ldots$, respectively, in the order created or in the order updated in CL 3.

CL 4: Repeat CL 1 through CL 3 until the remaining nodes have only outlinks or only inlinks.

After applying this algorithm, which consists of CL 1 through CL 4, to a given directed graph, the nodes are classified as $\mathrm{D}_{\mathrm{A}}, \mathrm{D}_{\mathrm{H}}, \mathrm{D}_{\mathrm{R}}, \mathrm{D}_{\mathrm{A}[1]}, \mathrm{D}_{\mathrm{H}[1]}, \mathrm{D}_{\mathrm{R}[1]} \ldots$. We refer to this method of node classification as the *node-clustering method*. We show the node-clustering method in the following two examples.

Example 2. (*Application to the graph in Fig. 1(a)*) We apply the node-clustering algorithm to the directed graph in Fig. 1(a). In order to perform the node-clustering, the two matrices, $\mathbf{T}_{\mathrm{A}}$ and $\mathbf{T}_{\mathrm{H}}$ in Eqn. (32), corresponding to Fig. 1(a), were generated; these are the matrices $\mathbf{T}_{\mathrm{A}_1}$ and $\mathbf{T}_{\mathrm{H}_1}$, respectively:

$$\mathbf{T}_{\mathrm{A}_1} = \begin{pmatrix} .586 & .305 & 256 & .305 & .586 & .343 \\ 1.62 & 3.11 & 1.80 & 2.18 & 1.62 & 1.19 \\ 3.69 & 4.88 & 8.45 & 8.12 & 3.69 & 7.85 \\ 3.06 & 4.12 & 5.66 & 5.89 & 3.06 & 5.02 \\ .586 & .305 & .256 & .305 & .586 & .343 \\ 2.73 & 1.79 & 4.34 & 3.98 & 2.73 & 4.67 \end{pmatrix}, \quad (34)$$

$$\mathbf{T}_{\mathrm{H}_1} = \begin{pmatrix} 3.66 & 3.14 & 2.17 & 2.82 & 3.22 & 2.17 \\ 6.25 & 7.30 & 3.30 & 6.93 & 7.14 & 3.30 \\ .336 & .256 & .567 & .319 & .335 & .567 \\ 3.60 & 4.44 & 2.63 & 4.68 & 4.36 & 2.63 \\ 3.81 & 4.23 & 2.56 & 4.02 & 4.33 & 2.56 \\ .336 & .256 & .567 & .319 & .335 & .567 \end{pmatrix}, \quad (35)$$

where all the entries in $\mathbf{T}_{\mathrm{A}_1}$ and $\mathbf{T}_{\mathrm{H}_1}$ are 10^2 times each. Using the node-clustering algorithm with respect to the two matrices $\mathbf{T}_{\mathrm{A}_1}$ and $\mathbf{T}_{\mathrm{H}_1}$ above, the results of node-clustering (CL 1 through CL 4) based on matrices $\mathbf{T}_{\mathrm{A}_1}$ and $\mathbf{T}_{\mathrm{H}_1}$ are presented in Table 2. As shown in Table 2, a total of 23 steps were required to complete the node-clustering algorithm, and the sets $\mathrm{D}^{(2)}_{\mathrm{A}}$, $\mathrm{D}^{(2)}_{\mathrm{H}}$, and $\mathrm{D}^{(1)}_{\mathrm{R}}$ were generated as follows:

$$\mathrm{D}^{(2)}_{\mathrm{A}} = \{n_3, n_4, n_6\},$$
$$\mathrm{D}^{(2)}_{\mathrm{H}} = \{n_2, n_4, n_5\},$$
$$\mathrm{D}^{(1)}_{\mathrm{R}} = \{n_4\}.$$

The process of Relation 23 in Table 2, $n_3 \circ\!\!\rightarrow n_5$, corresponds to the superfluous relation between the authority set and the hub set and terminates the initial step of node clustering. The remaining node, n_1, had only an outlink, so the node-clustering algorithm was completed by CL 4. Figure 7(a) shows the actual simulations of node-clustering for the directed graph in Fig. 1(a) when $k = 0$ and Fig. 7(b) presents the results of clustering. ♦

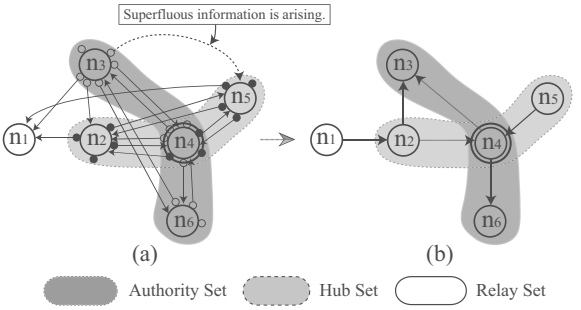


Fig. 7. Results of clustering for Fig. 1(a).

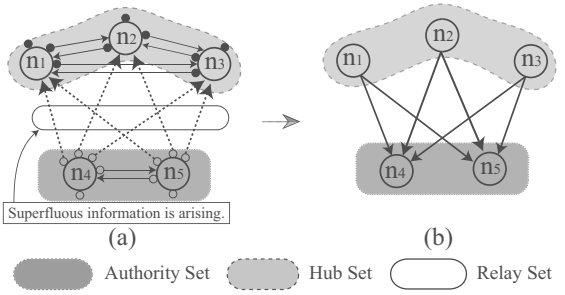


Fig. 8. Results of clustering for Fig. 1(b).

Example 3. (*Application to the graph in Fig. 1(b)*) We applied the node-clustering algorithm to the directed graph in Fig. 1(b). The graph in Fig. 1(b) is the so-called bipartite graph. Figure 8(a) presents the actual simulation of node clustering and Fig. 8(b) shows the results of performing node clustering for the graph in Fig. 1(b). As shown in Fig. 8(b), the nodes by the side of the outlinks generate the hub set, and the nodes by the side of the inlinks exactly generate the authority set. ♦

4.5. Method of generating a multistage directed graph. In order to obtain another expression of a

Table 2. Process of node clustering for Fig. 1(a).

No.	Relation	Set
1	$n_3 \circ$	
2	$n_3 \circ\!\!\rightarrow n_4$	
3	$n_3 \circ\!\!\rightarrow n_6$	
4	$n_2 \bullet$	
5	$n_2 \bullet\!\!\rightarrow n_5$	
6	$n_2 \bullet\!\!\rightarrow n_4$	
7	$n_2 \bullet\!\!\rightarrow n_1$	
8	$n_4 \circ$	
9	$n_4 \circ\!\!\rightarrow n_3$	$\mathrm{D}_{\mathrm{A}}^{(1)} = \{n_3, n_4\}$
10	$n_4 \circ\!\!\rightarrow n_6$	
11	$n_3 \circ\!\!\rightarrow n_2$	
12	$n_4 \bullet$	
13	$n_6 \circ$	
14	$n_4 \bullet\!\!\rightarrow n_2$	$\mathrm{D}_{\mathrm{H}}^{(1)} = \{n_2, n_4\}$ $\rightarrow \mathrm{D}_{\mathrm{R}}^{(1)} = \{n_4\}$
15	$n_4 \bullet\!\!\rightarrow n_5$	
16	$n_6 \circ\!\!\rightarrow n_3$	$\mathrm{D}_{\mathrm{A}}^{(2)} = \{n_3, n_4, n_6\}$
17	$n_5 \bullet$	
18	$n_5 \bullet\!\!\rightarrow n_2$	$\mathrm{D}_{\mathrm{H}}^{(2)} = \{n_2, n_4, n_5\}$
19	$n_4 \circ\!\!\rightarrow n_2$	
20	$n_5 \bullet\!\!\rightarrow n_4$	
21	$n_6 \circ\!\!\rightarrow n_4$	
22	$n_5 \bullet\!\!\rightarrow n_1$	
23	$n_3 \circ\!\!\rightarrow n_1$	
	$n_3 \circ\!\!\rightarrow n_5$	CL 1 is stopped.

directed graph, we now present a method for drawing a multistage directed graph. The initial step is to select the relations between nodes that belong to $\mathrm{D_A}$, $\mathrm{D_H}$, and $\mathrm{D_R}$, which are generated in the initial stage (corresponding to CL 1) in the node-clustering algorithm, and then to draw these relations between the corresponding nodes on the lowest-level plane; this is referred to as the bottom stage. As we mentioned in Section 4.2, each entry in $\{t_{\mathrm{A}}[i,j]\}$ and $\{t_{\mathrm{H}}[i,j]\}$ in Eqn. (32) is referred to as the degree of influence of the relation from n_i to n_j on a structure with respect to the authority and the hub, respectively, and if the entry in $\{t_{\mathrm{A}}[i,j]\}$ and $\{t_{\mathrm{H}}[i,j]\}$ is large, the degree of influence of the relation from n_i to n_j on the structure is also large. Thus, the relations of the nodes that belong to the sets of $\mathrm{D_A}$, $\mathrm{D_H}$, and $\mathrm{D_R}$, which were created during CL 1, have a substantial influence on the structure of the resulting directed graph. In the next step, we draw the relations between the nodes that belong to $\mathrm{D_{A[1]}}$, $\mathrm{D_{H[1]}}$, and $\mathrm{D_{R[1]}}$, which are generated in the first stage, on the plane that is located one step above the bottom stage. Repeating this process, a few more stages are constructed, and all of the nodes in the directed graph are divided into corresponding stages. The resulting graph is referred to as a *multistage directed graph*.

Then, we give the following definition and property.

Definition 7. (*Frame of a directed graph*) The relations of the nodes that belong to the sets of $\mathrm{D_A}$, $\mathrm{D_H}$, and $\mathrm{D_R}$ between the nodes on the bottom stage of a multistage directed graph are called the frame of the given directed graph.

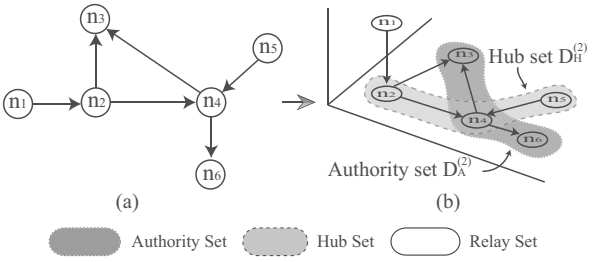


Fig. 9. Multistage directed graph in Fig. 1(a).

Property 7. The frame of a directed graph has a substantial influence on the structure of a given directed graph.

Figure 9 shows the relations between the nodes in the multistage directed graph for the directed graph in Fig. 1(a).

4.6. Verification of the node-clustering method. In this section, we verify the node-clustering method. For the first step, we prepare some directed graphs in which the structure of the directions among the nodes is very simple. We then apply the node-clustering method to these graphs and investigate whether clustering has occurred and if it has the characteristics of a hub set, an authority set, and a relay set. Since these methods are performed for directed graphs with simple structures, the value of k is assumed to be zero during PH1; this is to prevent it from having too much influence on the direction of the directed graph.

4.6.1. Verification for a directed graph having three nodes. We assume a graph that has three nodes, and we adopt all directed graphs in which the directions of the edge are entirely different. There are 11 distinct directed graphs that satisfy the conditions stated above. Figure 10 shows the results of performing node clustering for graphs of Type (A) through Type (K). In the following example, we present the actual calculations for a Type (I) graph as a typical example.

Example 4. *(Actual calculations for a Type (I) graph)* Here we present an example of node-clustering for a graph of Type (I). Two matrices, $\mathbf{T}_{\mathrm{A}}$ and $\mathbf{T}_{\mathrm{H}}$ in Eqn. (32), corresponding to Type (I) are generated as the matrices $\mathbf{T}_{\mathrm{A}_2}$ and $\mathbf{T}_{\mathrm{H}_2}$, respectively:

$$\mathbf{T}_{\mathrm{A}_2} = \begin{pmatrix} 0.278 & 0.0276 & 0.0276 \\ 0.0157 & 0.159 & 0.159 \\ 0.0157 & 0.159 & 0.159 \end{pmatrix}, \qquad (36)$$

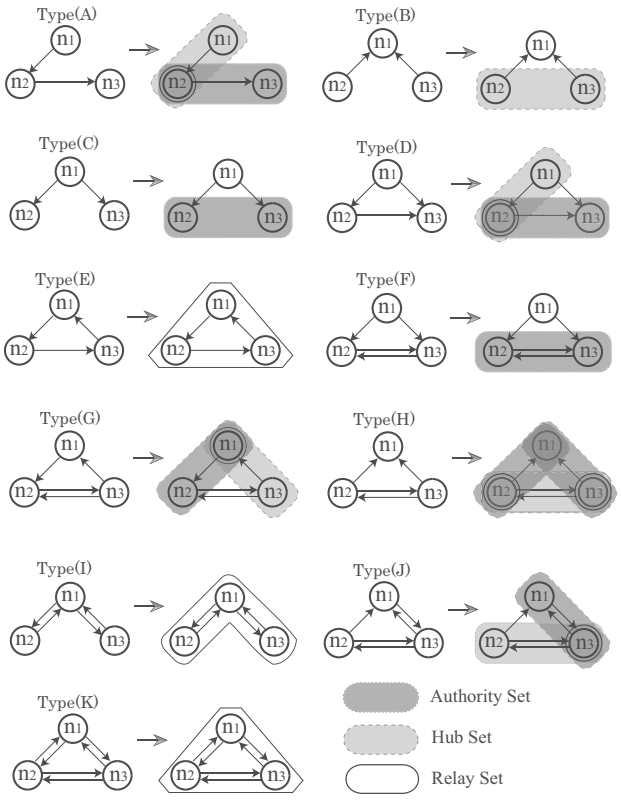


Fig. 10. Node clustering when $k = 0$.

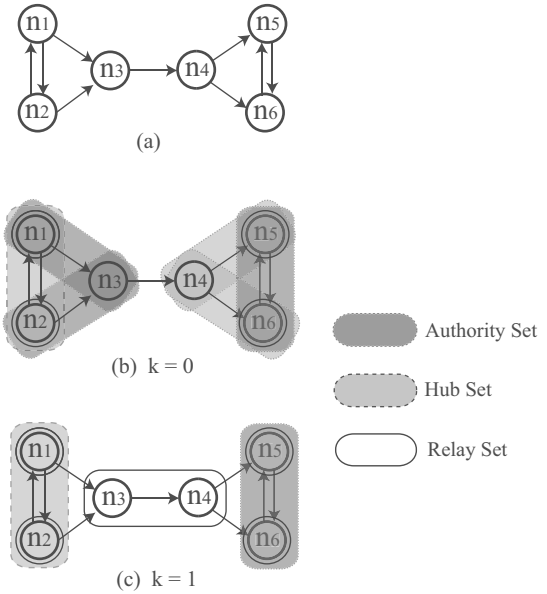


Fig. 11. Verification of the value of k.

Table 3. Process of node clustering for Type (I).

No.	Relation	Set
1	$n_1\circ$, $n_1\bullet$	
2	$n_2\circ$, $n_3\circ$	
	$n_2\bullet$, $n_3\bullet$	
	$n_2\circ\!\!\rightarrow n_3$	
	$n_3\circ\!\!\rightarrow n_2$	$\mathrm{D}_{\mathrm{A}}^{(1)} = \{n_2, n_3\}$
	$n_2\bullet\!\!\rightarrow n_3$	
	$n_3\bullet\!\!\rightarrow n_2$	$\mathrm{D}_{\mathrm{H}}^{(1)} = \{n_2, n_3\}$
		$\rightarrow \mathrm{D}_{\mathrm{R}}^{(1)} = \{n_2, n_3\}$
	$n_1\circ\!\!\rightarrow n_2$	
	$n_1\circ\!\!\rightarrow n_3$	
	$n_1\bullet\!\!\rightarrow n_2$	
	$n_1\bullet\!\!\rightarrow n_3$	
	$n_2\circ\!\!\rightarrow n_1$	$\mathrm{D}_{\mathrm{A}}^{(2)} = \{n_1, n_2\}$
		$\rightarrow \mathrm{D}_{\mathrm{A}}^{(3)} = \{n_1, n_2, n_3\}$
	$n_3\circ\!\!\rightarrow n_1$	
	$n_2\bullet\!\!\rightarrow n_1$	$\mathrm{D}_{\mathrm{H}}^{(2)} = \{n_1, n_2\}$
		$\rightarrow \mathrm{D}_{\mathrm{H}}^{(3)} = \{n_1, n_2, n_3\}$
		$\rightarrow \mathrm{D}_{\mathrm{R}}^{(3)} = \{n_1, n_2, n_3\}$
	$n_3\bullet\!\!\rightarrow n_1$	
	$n_3\bullet\!\!\rightarrow n_2$	

$$\mathbf{T}_{\mathrm{H}_2} = \begin{pmatrix} 0.278 & 0.0276 & 0.0276 \\ 0.0157 & 0.159 & 0.159 \\ 0.0157 & 0.159 & 0.159 \end{pmatrix}. \tag{37}$$

Table 3 shows the process of performing the node-clustering algorithm for a graph of Type (I). ♦

In Example 4, the matrices $\mathbf{T}_{\mathrm{A}_2}$ and $\mathbf{T}_{\mathrm{H}_2}$ are equal, which means that the structures directed by the authority and by the hub are the same. Such a phenomenon also occurs with the directed graphs of Type (E) and Type (K). Thus we have the following property.

Property 8. In general, if $\mathbf{T}_{\mathrm{A}} = \mathbf{T}_{\mathrm{H}}$ for a given directed graph, then all nodes are clustered as the one relay set.

When the node-clustering method is applied to the directed graphs of Type (A) through Type (K), the nodes are clustered into an authority set, a hub set, and a relay set, exactly in accordance with the directions of the edges of the graph. We thus conclude that the node-clustering method functions properly for the given directed graph.

4.6.2. Verification for different values of *k*. We now consider the influence of the value of k. We begin with the directed graph in Fig. 11(a) in order to more easily see the influence of different values of k. Node clustering for Fig. 11(a) when $k = 0$ appears to be performed by two groups, $\{n_1, n_2, n_3\}$ and $\{n_4, n_5, n_6\}$, as shown in Fig. 11(b). But if we consider the entire directed graph in Fig. 11(a), the set $\{n_1, n_2, n_3\}$ can be regarded as a hub set and the set $\{n_4, n_5, n_6\}$ can be regarded as an authority set. In order to better understand the graph, we will increase the

value of k. We note that as the value of k increases, the degrees of relation between the nodes also increases. We applied the node-clustering algorithm to Fig. 11(a) with $k = 1$ in order to change the relations between the nodes. For the case $k = 1$, which is the maximum k value, the results are shown in Fig. 11(c). From Fig. 11(c), we see that the nodes are clustered into the authority, hub, and relay sets, as expected. Therefore, we have the following property for the node-clustering method.

Property 9. When applying the node-clustering method to a directed graph, increasing the value of k is sufficient for joining all the nodes in a single cluster.

As a verification of Property 9, we give the following example.

Example 5. *(Verification of Property 9)* Figure 12 presents the node-clustering algorithm for the directed graph in Fig. 1(c). When $k = 0$, which results in the least influence of the distribution of the directed edges, the node clustering did not perform well (see Fig. 12(a)). The association of the authority and hub sets was confused by the directed edges that point to the hub set from the authority set. With an increased k ($k = 0.35$), the node clustering was completed with a connection between the authority and hub sets (see Fig. 12(f)). ♦

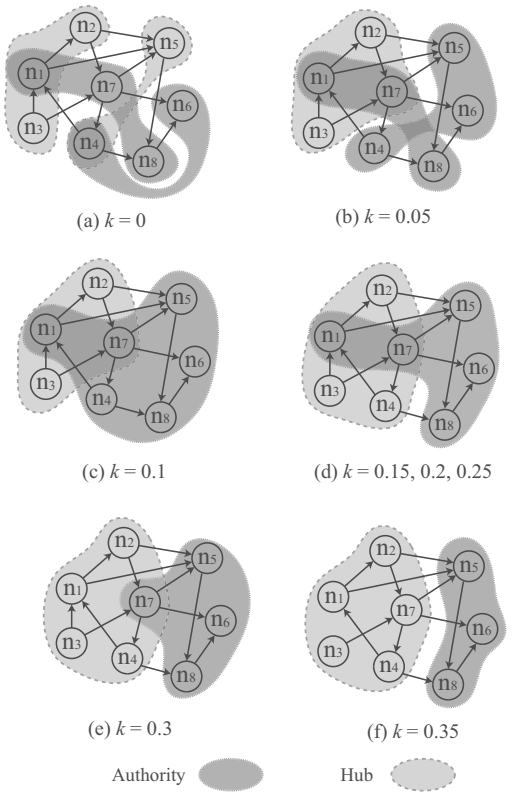


Fig. 12. Simulations with different values of k.

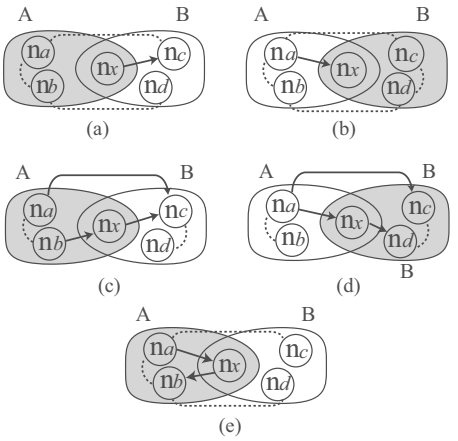


Fig. 13. How to treat the relay node.

4.7. Method of generating an improved directed graph. We now present a method that generates an improved directed graph, a simplified version of a given directed graph. An improvement should maintain the relations between $\mathrm{D_A}$, $\mathrm{D_H}$, and $\mathrm{D_R}$ at the corresponding stages of the multistage directed graph, and it is clear that it should be simpler than the original graph. One problem that arises is how to treat the relay set. The relay nodes belong to both the authority set and the hub set. Thus, if the relay nodes can be absorbed into either the authority or the hub sets, it will simplify the graph. We give the following conditions for how to treat the relay set.

Condition 3. Let $n_x \in \mathrm{D_R}$, either of A or B be an authority set and the other be hub set, and $\mathrm{A} \cap \mathrm{B} = \mathrm{D_R}$. The treatment of the relay nodes is as follows:

(a) If the relay set $\mathrm{D_R} = \mathrm{A} = \mathrm{B}$, then $\mathrm{D_R}$ is drawn as itself (corresponds to Type (E), Type (I), and Type (K) in Section 4.6).

(b) If n_x has only outlinks to the nodes belonging to B, then n_x is absorbed into A (see Fig. 13(a)).

(c) If n_x has only inlinks from nodes belonging to A, n_x is absorbed into B (see Fig. 13(b)).

(d) If n_x has both inlinks and outlinks, then treat it as follows:

(i) if all nodes belonging to A have outlinks to the nodes belonging to $\mathrm{D_R}$ and $\mathrm{B} - \mathrm{D_R}$, then n_x is absorbed into A (see Fig. 13(c));

(ii) if all nodes belonging to B have inlinks from the nodes belonging to $\mathrm{D_R}$ and $\mathrm{A} - \mathrm{D_R}$, then n_x is absorbed into B (see Fig. 13(d));

(iii) for cases in which (i) and (ii) are not satisfied between A and B, if all nodes belonging to A have both outlinks to n_x and inlinks from n_x, then n_x is absorbed into A (see Fig. 13(e)).

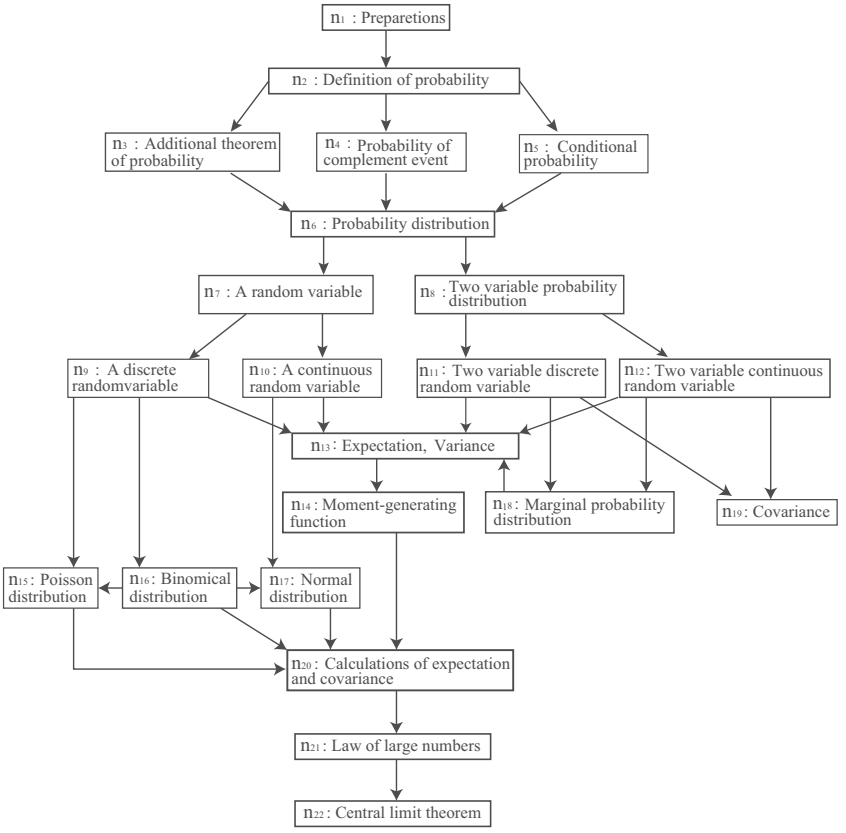


Fig. 14. Relations between topics from a text syllabus.

5. Application of the node-clustering method

Teachers usually prepare a syllabus that presents the students the planned subjects of the upcoming lectures. A syllabus is an important tool for presenting the contents of the course, but the standard form of a syllabus is text, which makes it difficult for students to grasp how the topics are related. In order to solve this problem, we offer a graphic syllabus (Nilson, 2007). It may be expressed in various ways, such as by the flow between the topics or by the teacher's individual abstract overview. In this study, we consider the flow between related topics and generate a graphic syllabus as an application of the node-clustering method.

5.1. Results of the application. Based on the text syllabus for a lecture in statistics, Fig. 14 shows the flow of the relations between topics. Applying the node-clustering method to this flow, we obtained a multistage directed graph (the value of k in Eqn. (14) was assumed to be 0.5). As shown in Fig. 15, this multistage directed graph required three stages of construction. The process of node clustering required 88 steps in the initial stage of clustering, and 38 more in the next stage. Table 4 shows the results of generating each of the sets at the corresponding stage. Figure 16(a) is an improved directed graph, obtained by using the method presented in Section 4.7. Each of the relay nodes was absorbed into a set in accordance with Condition 3 as follows:

Table 4. Generated sets corresponding to each stage.

Stage	Bottom stage	First stage
authority	$\mathrm{D}_{\mathrm{A}}^{(4)}$	$\mathrm{D}_{\mathrm{A}[1]}^{(1)}$, $\mathrm{D}_{\mathrm{A}[1]}^{(4)}$
hub	$\mathrm{D}_{\mathrm{H}}^{(1)}$, $\mathrm{D}_{\mathrm{H}}^{(3)}$	$\mathrm{D}_{\mathrm{H}[1]}^{(1)}$, $\mathrm{D}_{\mathrm{H}[1]}^{(2)}$, $\mathrm{D}_{\mathrm{H}[1]}^{(3)}$
relay	$\mathrm{D}_{\mathrm{R}}^{(1)}$, $\mathrm{D}_{\mathrm{R}}^{(2)}$	$\mathrm{D}_{\mathrm{R}[1]}^{(1)}$, $\mathrm{D}_{\mathrm{R}[1]}^{(2)}$

$$n_{18} \in \mathrm{D}_{\mathrm{H}}^{(3)}, \quad n_{17} \in \mathrm{D}_{\mathrm{H}[1]}^{(2)}, \quad n_7 \in \mathrm{D}_{\mathrm{A}[1]}^{(1)}. \tag{38}$$

Figure 16(a) presents the results of the graphic syllabus corresponding to the text syllabus, and Fig. 16(b) presents the frame of the multistage directed graph in Fig. 15.

5.1.1. Merits of a graphic syllabus. A graphic syllabus facilitates the following:

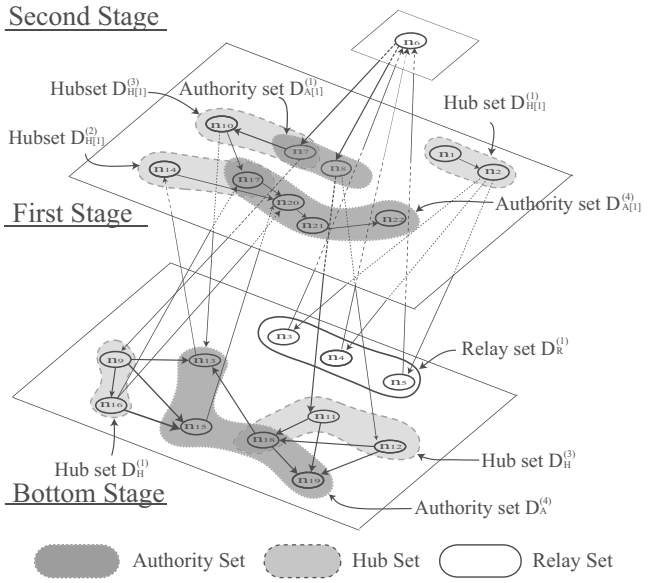


Fig. 15. Multistage directed graph using the node-clustering method.

(5a) The way in which the topics (teaching units) connect with the subjects as a whole is more easily comprehended by the students.

(5b) The key points are highlighted.

As for (5a), consider Fig. 16(b), the frame of the multistage directed graph. Note that the shape of this graph is a perfect diamond. Teaching units in $D^{(4)}_{A}$ at the bottom stage are the most significant for students to learn, and the teaching units in $D^{(1)}_{H}$ and $D^{(3)}_{H}$ are the most important as introductory units. The teaching units in $D^{(1)}_{R}$ are very important connectors, since even if students are able to understand the topics in $D^{(1)}_{H}$ and $D^{(3)}_{H}$, if they do not understand the topics in $D^{(1)}_{R}$, they will be unable to understand the topics in $D^{(4)}_{A}$. So when teaching units in $D^{(1)}_{R}$, it is important that students learn and thoroughly understand the teaching unit in $D^{(1)}_{R}$. From the flow in Fig. 14, we see that the final goal is to teach the central limit theorem. But as shown in the frame of the directed graph, the main task is to teach the concepts of expectation, variance, and covariance.

6. Conclusions

Our analysis of the structure of a directed graph can be applied to various topics in which there are given relations between nodes. In the present paper, we considered the flow between related topics in statistics and generate a graphic syllabus as an application of the node-clustering method.

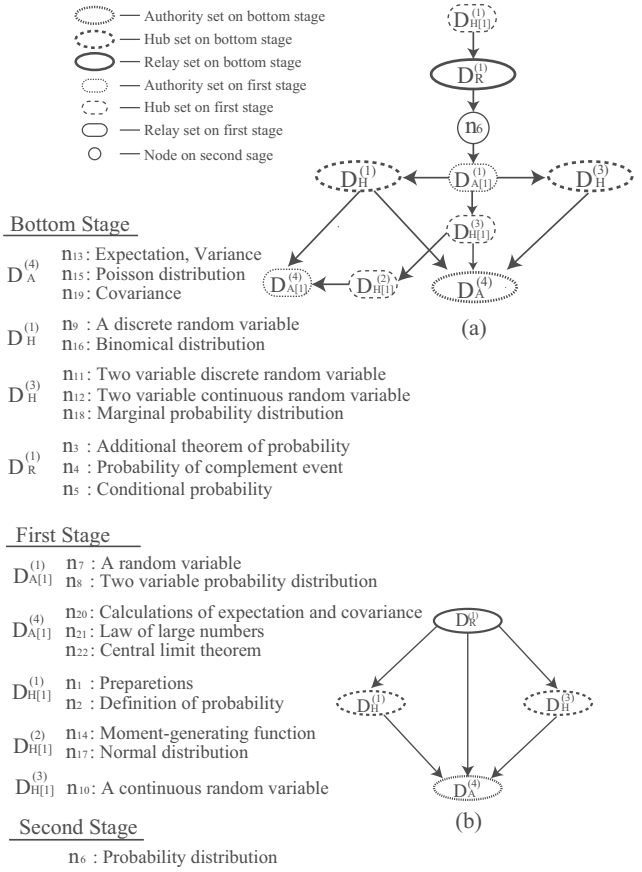


Fig. 16. Graphic syllabus using the node-clustering method.

Generating the authority, hub, and relay sets is one way to simplify a directed graph. We note that the method we presented in the present paper simplifies the graph by considering the distribution of the directed edges from the viewpoints of both the output and the input. Thus, we can even simplify a bipartite graph by using the authority and hub sets (see Fig. 8). We also presented a simplification method that considers the treatment of the relay nodes (see Section 4.7) and presented a method for generating an improved directed graph.

We performed a fundamental study of node clustering for a small directed graph, but further studies will be required to apply node clustering to very large directed graphs, such as the World Wide Web. As a first step, we began studying how to determine the value of k in PH1 for the structure of a given directed graph.

Acknowledgment

The authors wish to thank the anonymous reviewers for their careful reading and helpful suggestions.

References

Amy, N. and Carl, D. (2005). A survey of eigenvector methods for web information retrieval, *SIAM Review* **47**(1): 135–161.

Amy, N. and Carl, D. (2008). *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ.

Aracena, J. and Gomez, L. (2013). Limit cycles and update digraphs in Boolean networks, *Discrete Applied Mathematics* **161**(1–2): 217–243.

Balakrishnan, V.K. (1997). *Schaum's Outline of Theory and Problems of Graph Theory*, McGraw-Hill, New York, NY.

Berge, C. (2001). *The Theory of Graphs*, Dover Pubns, New York, NY.

Berman, A. and Plemmons, R. (1979). *Nonnegtive Matrices in the Mathematical Science*, Academic Press, New York.

Berry, M., Drmac, Z. and Jessup, E. (1999). Matrices, vector space, and information retrieval, *SIAM Review* **41**(2): 335–362.

Hofuku, I. and Oshima, K. (2006). Rankings schemes for various aspects based on Perron–Frobenius theorem, *Information* **9**(1): 37–52.

Hofuku, I. and Oshima, K. (2008). A controlled absolute ranking method applied to an exam of multiplex choice form, *International Journal of Pure and Applied Mathematics* **47**(2): 267–280.

Hofuku, I. and Oshima, K. (2010a). A mathematical structure of processes for generating rankings through the use of nonnegative irreducible matrices, *Applied Mathematics and Information Science* **4**(1): 125–139.

Hofuku, I. and Oshima, K. (2012). A new ranking model using the power method, *Applied Mathematics and Information Science* **6**(1): 75–84.

Hofuku, I., Yokoi, T. and Oshima, K. (2010b). Measures to represent the properties of nodes in a directed graph, *Information* **13**(3): 537–549.

Lancaster, P. and Tismenetsky, M. (1985). *The Theory of Matrices*, Academic Press, New York, NY.

Ligęza, A. and Kościelny, J.M. (2008). A new approach to multiple fault diagnosis: A combination of diagnostic matrices, graphs, algebraic and rule-based models. The case of two-layer models, *International Journal of Applied Mathematics and Computer Science* **18**(4): 465–476, DOI: 10.2478/v10006-008-0041-8.

Nilson, L. (2007). *The Graphic Syllabus and the Outcomes Map: Communicating Your Course*, Jossey-Bass, San Francisco, CA.

Ortega, J. (1990). *Numerical Analysis, A Second Course*, SIAM, Philadelphia, PA.

Prelim, J. and Demongeot, E. (2013). On the number of update digraphs and its relation with the feedback arc sets and tournaments, *Discrete Applied Mathematics* **161**(10–11): 1345–1355.

Yang, F., Shah, S. and Xiao, D. (2012). Signed directed graph based modeling and its validation from process knowledge and process data, *International Journal of Applied Mathematics and Computer Science* **22**(1): 41–53, DOI: 10.2478/v10006-012-0003-z.

Yokoi, T. and Hofuku, I. (2010). The keyword extraction with the ranking method using ANP, *Information* **13**(3(B)): 1065–1073.

Ichiro Hofuku is a professor at the Tokyo Metropolitan College of Industrial Technology. He received a Dr.Sci. degree in information science from the Tokyo University of Science in 1999. His research area is applied mathematics, particularly various kinds of mathematical models. At present, he is engaged in investigating a technique for controlling the rank inversion phenomenon in a ranking generation process. Moreover, recently, he has started to design a new mathematical system by analyzing the structures of both network structures and information retrieval systems. He is a member of the JSIAM, ORSJ and MPS.

Kunio Oshima is a professor in the Graduate School of Management Science and the School of Management Science at the Tokyo University of Science. He received a Ph.D. degree in mathematics from the University of Houston in 1979. His current interests include applications of non-negative matrices to definite ranking methods and management science. He is a member of the SIAM, JSIAM and IPSJ.

Received: 25 December 2012
Revised: 16 June 2013
Re-revised: 7 August 2013