Computing the Existence of Paths within a Digraph by Warshall's Algorithm

Mya Mya Thin[1] and Myo Nilar[2]

Abstract

Warshall's algorithm is an efficient method to determine the transitive closure of a digraph or all paths in a digraph by using the adjacency matrix. In this paper, we firstly introduce the basic concepts of graph. Then, we present the adjacency matrix of a digraph. Next we study the Warshall's algorithm for finding the reachability matrix of a digraph. Finally, we calculate all paths in a digraph by using Warshall's algorithm.

Keywords: Adjacency matrix, Reachability matrix, Warshall's algorithm.

Introduction

Warshall's algorithm is a powerful graph algorithm used to find the transitive closure of a digraph. It was developed by Stephen Warshall in 1962 and has wide applications in various field such as computer science, operations research, and social network analysis. By iteratively considering all pairs of vertices in the digraph, Warshall's algorithm efficiently determines whether or not there exists a path between any two vertices in the digraph. This transitive closure reveals valuable information about the connectivity of the graph and can be used in various graph related problems. This paper is divided into four sections. First section introduces the basic concepts of graph theory with illustrative examples. Second section presents the adjacency matrix of a digrph. Third section discusses the Warshall's algorithm and construction of a reachability matrix. Fourth section apply the Warshall's algorithm to calculate all paths in a digraph.

Basic Concepts of Graphs

Graphs are represented graphically by drawing a dot or circle for every vertex, and drawing an arc between two vertices if they are connected by an edge. For graph notations and technologies used in this paper, we refer to [2], [4] and we recall some of them.

A **graph** G is a finite nonempty set of objects called **vertices** (the singular is vertex) together with a (possibly empty) set of unordered pairs of distinct vertices of G called **edges**. The vertex set of G is denoted by $V(G)$, while the edge set is denoted by $E(G)$. A graph having no edges is called a **null graph**. A **digraph** D is a finite nonempty set of objects called **vertices** together with a set of ordered pairs of vertices called **arcs**. The vertex set of D is denoted by $V(D)$, while the arc set is denoted by $E(D)$. Throughout this paper, we consider only digraphs.

[1] Lecturer, Department of Mathematics, Hinthada University
[2] Professor and Head, Dr., Department of Mathematics, Hinthada University

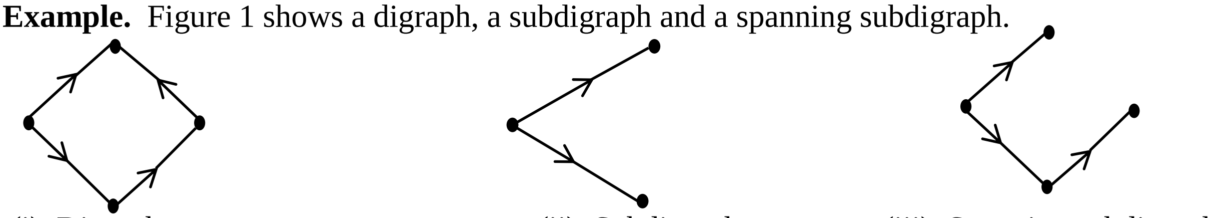
The **order** of a digraph D is the cardinality of its vertex set and is denoted by $n(D)$ or simply n; n_D is also used when dealing with many digraphs at the same time. The **size** of a digraph D, denoted by $m(D)$ or m is the cardinality of its arc set. If $a=(u,v)=uv$ is an arc of a digraph D (a is said to join u to v), then a is said to be **incident from** u and **incident to** v. Consequently, u is **adjacent to** v, whilst v is **adjacent from** u. For an arc uv, the first vertex u is also known as its **tail** and v is known as the arc uv's **head**. A **loop** is an arc whose head and tail coincide. A digraph in which parallel arcs and loops are allowed, is defined to be a **directed pseudograph**. A directed pseudograph in which no loops occur is known as a **directed multigraph**.

Subdigraphs, Deletion and Contraction

In this section, we give the definitions of subdigraphs, deletion, contraction and illustrated examples.

Definitions. A digraph H is a **subdigraph** of a digraph D if $V(H)\subseteq V(D)$, $E(H)\subseteq E(D)$ and every arc in $E(H)$ has both end-vertices in $V(H)$. If $V(H)=V(D)$, then H is a **spanning subdigraph** of D.

Example. Figure 1 shows a digraph, a subdigraph and a spanning subdigraph.



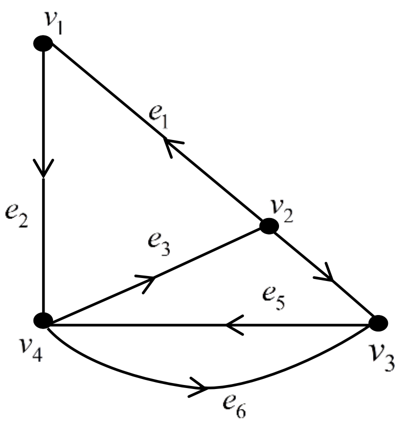
(i) Digraph (ii) Subdigraph (iii) Spanning subdigraph

Figure 1.

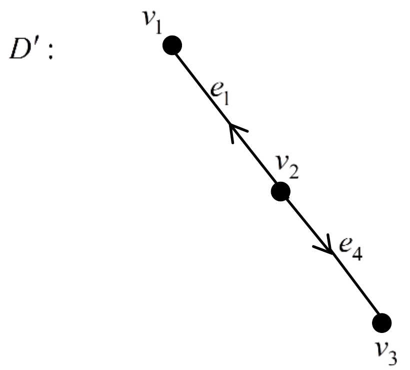
Definitions. If every arc of $E(D)$ with both end-vertices in $V(H)$ is in $E(H)$, we say that H is induced by $U=V(H)$ and call H an **induced** subdigraph or **vertex induced** subdigraph of D. For a subset $A'\subseteq E(D)$, the subdigraph **arc-induced** by A' is the digraph $D\langle A'\rangle=(V',A')$, where V' is the subset of the vertices in $V(D)$ that are incident with the arcs in A'. A **vertex-deleted subdigraph** of a digraph D is obtained by deleting/ removing one or more vertices from $V(D)$ and all the arcs in $E(D)$ that are incident with the removed vertex/ vertices. **Arc-deleted subdigraphs** of a digraph D are obtained by simply removing arcs from D. If H' is an arc-deleted subdigraph, then $V(H')=V(D)$ and $E(H')\subset E(D)$.

Example. Consider the digraph D with the vertex set $V(D)=\{v_1,v_2,v_3,v_4\}$ and the arc set $E(D)=\{e_1,e_2,e_3,e_4,e_5,e_6\}$ shown in Figure 2.

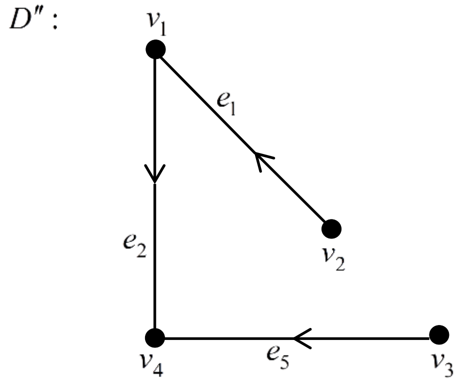
$D:$



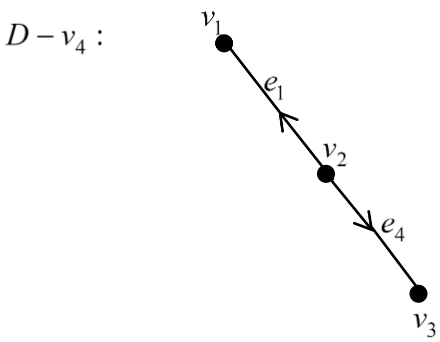
The subdigraph D' of D induced by $U = \{v_1, v_2, v_3\}$ is given below.



The subdigraph D'' of D induced by $A' = \{e_1, e_2, e_5\}$ is given below.



If we delete the vertex v_4 from D, we get the following subdigraph of D.



If we delete the arc e_4 from D, we get the following subdigraph of D.

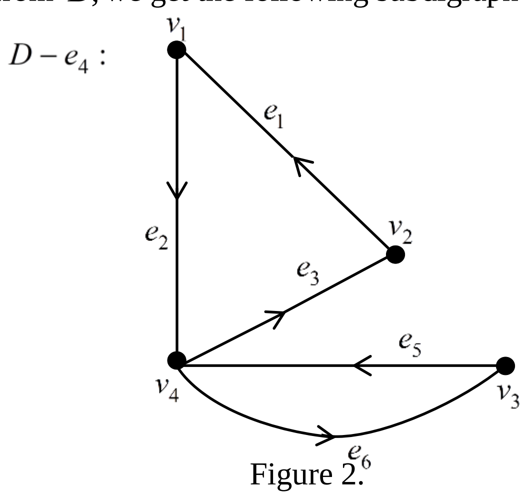


Figure 2.

Definition (Contraction in a digraph). The **contraction** of an arc e with the endpoints u and v is the replacement of u, v and e with a single vertex, say v', whose incident arcs are all the arcs other than e that were incident with u and v. D/e denote the arc contraction of the arc e in the digraph D.

Example. Figure 3 shows a digraph D with the vertex set $V(D) = \{v_1, v_2, v_3, v_4, v_5\}$ and the arc set $E(D) = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$. D/e_6 is depicted by D'.

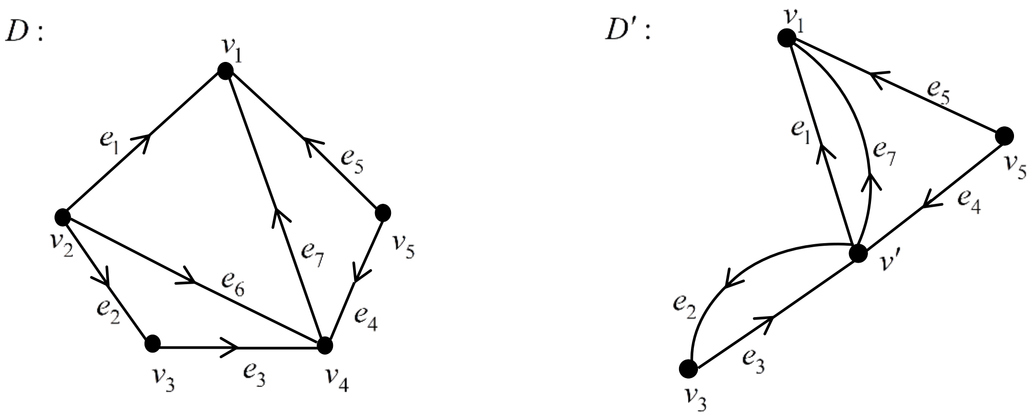


Figure 3.

The cardinalities of the vertex and arc sets are $|V(D')| = 4$ and $|E(D')| = 6$.

Adjacency Matrix

In this section, we present the adjacency matrix of a digraph. This concept extends to directed pseudograph.

Definition. Let $D = (V, E)$ be a digraph, where $V = \{v_1, v_2, \ldots, v_n\}$. The **adjacency matrix** of a digraph D is the $n \times n$ matrix $A = (a_{ij})$, the entry in the i-th row and j-th column is defined by

$$a_{ij} = \begin{cases} 1 & \text{if } v_i v_j \in E, \\ 0 & \text{if } v_i v_j \notin E. \end{cases}$$

Example. Consider the digraph D with the vertex set $V(D) = \{v_1, v_2, v_3, v_4, v_5\}$ and the arc set $E(D) = \{(v_2, v_1), (v_2, v_3), (v_4, v_3), (v_3, v_4), (v_2, v_4), (v_5, v_4), (v_5, v_1)\} = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$ shown in Figure 4.

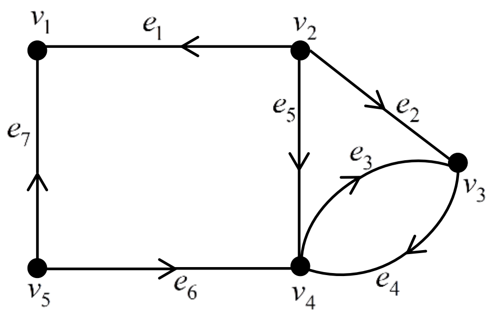


Figure 4.

The adjacency matrix A of a digraph D is

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Walks, Trails, Paths and Cycles

Let $D=(V,E)$ be a directed pseudograph for the discussion on the following definitions: walks, trails, paths and cycles.

Definitions. A **walk** W in D is an alternating sequence of vertices and arcs in D. We write $W = x_1a_1x_2a_2\ldots a_{k-1}x_k$ such that x_i is the tail of arc a_i and x_{i+1} is the head, where $i = 1, 2, \ldots,$ $k-1$. A (directed) walk W is **closed** if the initial vertex x_1 in the sequence of W coincides with the terminal vertex x_k in W, i.e., $x_1 = x_k$. A (directed) **trail** of D is a (directed) walk in which all arcs are distinct. A (directed) **path** of a directed pseudograph D is a (directed) walk of D in which no vertices or arcs are repeated. A (directed) **cycle** of a directed pseudograph D is a closed directed path in D. When we speak of a (directed) $u-v$ walk/ path, then we are referring to a directed walk/ path with the initial vertex u and the terminal vertex v. The **length** of a (directed) path, (directed) cycle or (directed) trail in D is also defined as the number of arcs present. A (directed) cycle of length k is referred to as a (directed) ***k*-cycle**.

Example. Consider the digraph D shown in Figure 5. The possible (directed) v_2-v_1 walks and (directed) v_2-v_1 trails are given in Table 1.

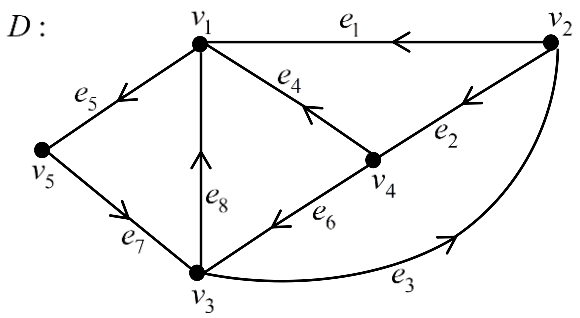


Figure 5.

Table 1.

Possible (directed) v_2-v_1 walks	Possible (directed) v_2-v_1 trails
$W_1:\ v_2v_1$	$T_1:\ v_2v_1$
$W_2:\ v_2v_4v_1$	$T_2:\ v_2v_4v_1$
$W_3:\ v_2v_4v_1v_5v_3v_1$	$T_3:\ v_2v_4v_1v_5v_3v_1$
$W_4:\ v_2v_4v_3v_1$	$T_4: v_2v_4v_3v_1$
$W_5:\ v_2v_1v_5v_3v_1$	$T_4:\ v_2v_1v_5v_3v_1$
$W_6:\ v_2v_1v_5v_3v_2v_1$	

The cycle $C_5: v_4v_1v_5v_3v_2v_4$ is highlighted on the digraph below:

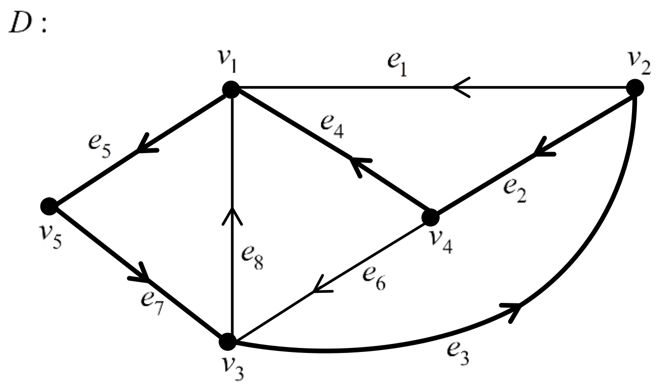


Figure 5.

Looking at the above, we can see that a directed path is contained in a directed cycle. One such path is $P_4: v_4v_1v_5v_3v_2$. Some $v_2 - v_1$ paths are listed in Table 2.

Table 2.

$v_2 - v_1$ paths
$P_1: v_2v_1$
$P_2: v_2v_4v_1$

The adjacency matrix of D is given by

$$A(D) = A = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

We now present the following two theorems on the existence of a directed path in a digraph. The proofs of these theorems can be seen in [1].

Theorem. Let D be a digraph and u and v be two distinct vertices such that $u, v \in V(D)$. If D has a closed (directed) $u - v$ walk, then W contains a cycle C through u such that $E(C) \subseteq E(W)$.

Theorem. Let D be a digraph and u and v be two distinct vertices such that $u, v \in V(D)$. If D has a (directed) $u - v$ walk, then D contains a (directed) $u - v$ path.

We also define the next two definitions.

Definition. The length of the shortest (directed) path between two vertices u and v in a digraph D is known as the **distance** between u and v. We write $d(u, v)$.

Definition. A digraph D is **acyclic** if it does not contain any cycles.

Reachability: Warshall's Algorithm

In this section, we study Warshall's algorithm for finding the reachability matrix of a digraph [3].

Definition. A vertex v is **reachable** from a vertex u in a digraph / directed pseudograph D if D has a $u-v$ walk.

Definition. The **reachability matrix** of a digraph (or a directed pseudograph) D is an $n \times n$ matrix $R(D)$ where r_{ij}, the entry in the i^{th} row and the j^{th} column of $R(D)$, is defined by

$$r_{ij} = \begin{cases} 1 & \text{if } j \text{ is reachable from } i, \\ 0 & \text{if } j \text{ is not reachable from } i. \end{cases}$$

Boolean Addition in Terms of Vectors

Let $\vec{u}$ and $\vec{v}$ be $1 \times n$ vectors in R^n, where each entry is either 0 or 1. Define the following addition between vectors:

$\vec{u} + \vec{v} = [u_1 \quad u_2 \quad \dots \quad u_n] + [v_1 \quad v_2 \quad \dots \quad v_n]$; where $u_i = 0$ or 1 for $i = 1, 2, \dots, n$

$v_j = 0$ or 1 for $j = 1, 2, \dots, n$.

$$= [u_1 + v_1 \quad u_2 + v_2 \quad \dots \quad u_n + v_n]$$

We add the components as follows for $i = 1, 2, \dots, n$:

$$u_i + v_i = \begin{cases} 0 + 1 = 1 & \text{if } u_i = 0 \text{ and } v_i = 1 \\ 1 + 0 = 1 & \text{if } u_i = 1 \text{ and } v_i = 0 \\ 0 + 0 = 0 & \text{if } u_i = 0 \text{ and } v_i = 0 \\ 1 + 1 = 1 & \text{if } u_i = 1 \text{ and } v_i = 1. \end{cases}$$

We use the above vector addition for the row operations in the construction of the reachability matrix.

Construction of a Reachability Matrix in Warshall's Algorithm

Let D be a digraph and $V(D) = \{1, 2, \dots, n\}$. Let Q be an $n \times n$ matrix with the entries

$$q_{ij} = \begin{cases} 1 & \text{if } j \text{ is reachable from } i \\ 0 & \text{if it is not known whether } j \text{ is reachable from } i \end{cases}$$

for $i, j = 1, 2, \dots, n$.

Then, $A(D)$ will give the initial set of vertices that are reachable from each other.

Therefore, $Q_0 = A(D)$.

Take a vertex j and keep it fixed.

(1) Consider row j of Q. If k is reachable from j, then $q_{jk} = 1$. Otherwise, $q_{jk} = 0$.

(2) Considering the j^{th} column of Q. If j is reachable from i, then $q_{ij} = 1$. Otherwise, $q_{ij} = 0$.

(3) From (1) and (2), if j is reachable from i and k is reachable from j, then k is reachable from i, since we have a $i-j$ walk followed by a $j-k$ walk which is a $i-k$ walk. Therefore, replace the entry q_{ij} by 1 in Q if $q_{ij} = 0$.

In step (3), we are performing Boolean addition on the rows of Q. We add $\text{Row}_j(Q) = R_j$ to R $\text{Row}_i(Q) = R_i$ when the entry $q_{ij} = 1$.

The following algorithm can be used to find the reachability matrix R of a digraph D.

Warshall's Algorithm

Let $D = (V, E)$ be a digraph, where $V(D) = \{1, 2, ..., n\}$ and $A(D)$ be the adjacency matrix of D. Let R_i denote row i and Col_j denote column j. The steps of the Warshall's Algorithm are as follows.

Step 1. Set $Q_0 = A(D)$, initialization.

Step 2. For $i = 1$ to n do the following.

Add R_i to every row of Q_{i-1} in which 1 is the entry in the i^{th} column of Q_{i-1}.

Step 3. Set $R = Q_n$.

This algorithm can be used in the following example.

Example. Consider the directed pseudograph D shown in Figure 6.

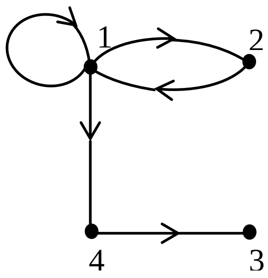


Figure 6.

The adjacency matrix is

$$Q_0 = A(D) = A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

From Warshall's algorithm, column 1 has 1's in row 2; therefore, add R_1 to R_2.

$$\xrightarrow{R_2 \leftarrow R_2 + R_1} \quad Q_1 = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

For Q_2, rows 1 has 1 in their 2nd columns, therefore

$$\xrightarrow{R_1 \leftarrow R_1 + R_2}$$

$$Q_2 = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

For Q_3, row 4 of Q_2 has 1 in their 3rd columns, therefore

$$\xrightarrow{R_4 \leftarrow R_4 + R_3}$$

$$Q_3 = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

For Q_4, rows 1 and 2 of Q_3 has 1 in their 4th columns. Thus

$$\xrightarrow{R_1 \leftarrow R_1 + R_4, R_2 \leftarrow R_2 + R_4}$$

$$Q_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Hence, $R = Q_4$.

This shows that the vertices 1, 2, 3, and 4 are reachable from 1 and 2.

The vertices 3 is reachable from 4.

We can see all possible paths in the following Figures 7.

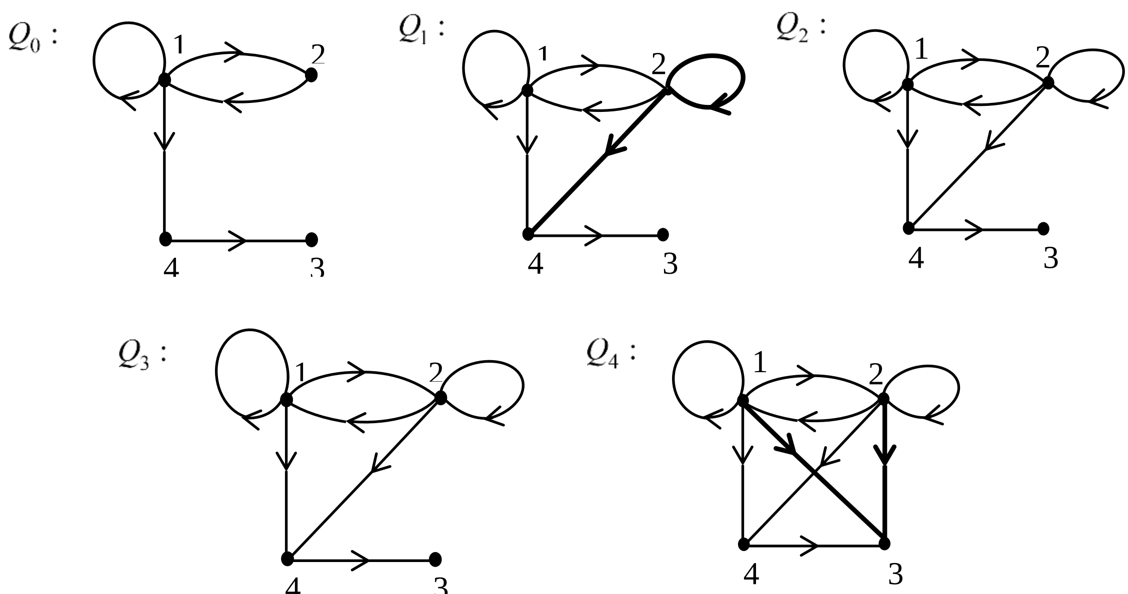


Figure 7.

Conclusion

Graphs are used to model many problems of the real world in the various fields. The main advantage of the presented approach is a Warshall's algorithm to find the reachability matrix. This algorithm provides a simple and efficient solution to the problem of finding the existence of paths in a digraph. Converting the adjacency matrix A into the reachability matrix R by the Warshall's algorithm, we obtain all the possible paths through the digraph between each pair of vertices.

Acknowledgements

I would like to express our gratitude to Dr Theingi Shwe, (Rector, Hninthada University), Dr Yee Yee Than (Pro-Rector, Hninthada University), and Dr Aye Lwin (Pro-Rector, Hninthada University) for their permission to summit this research paper. Special thanks are also due to Dr Myo Nilar, Professor and Head, Department of Mathematics, Hinthada University, for her invaluable advice, suggestions and guidance.

References

G. Chartrand, L. Lesniak, P. Zhang, "*Graphs and Digraphs*" , Sixth Edition, Chapman & Hall Book / CRC Press, New York, (2016).

J. B. Jensen, G. Z. Gutin, "*Digraphs*: *Theory*, *Algorithms and Applications*", Second Edition, Springer-Verlag, London, (2009).

K. Ruohonen, "*Lecture notes on Graph Theory*", Tampere University of Technology, Finland, (2008).

S. S. Ray, "*Graph Theory with Algorithms and its Applications*", 2013th Edition, Springer, New York, (2013).