



Mathematical Foundations for Data Science

Pilani Campus MFDS Team



DSECL ZC416, MFDS

Lecture No.12

Agenda

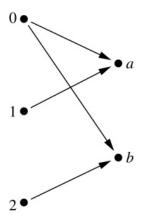
- Relations and its types
- n- ary relations
 - Join
 - Select
 - Projection
- Representation of Relations
- Closures of relations
 - Warshalls algorithm
 - Operation count

Binary Relations

Definition: A binary relation R from a set A to a set B is a subset $R \subseteq A \times B$.

Example:

- Let $A = \{0,1,2\}$ and $B = \{a,b\}$
- $-\{(0, a), (0, b), (1,a), (2, b)\}\$ is a relation from A to B.
- We can represent relations from a set A to a set B graphically or using a table:



R	а	b
0	×	×
1	×	
2		×

Relations are more general than functions. A function is a relation where exactly one element of *B* is related to each element of *A*.

Binary Relation on a Set

Definition: A binary relation R on a set A is a subset of $A \times A$ or a relation from A to A.

Example:

- -Suppose that $A = \{a,b,c\}$. Then $R = \{(a,a),(a,b),(a,c)\}$ is a relation on A.
- -Let $A = \{1, 2, 3, 4\}$. The ordered pairs in the relation
- R = $\{(a,b) \mid a \text{ divides } b\}$ are (1,1), (1, 2), (1,3), (1, 4), (2, 2), (2, 4), (3, 3), and <math>(4, 4).

Question: How many relations are there on a set A? $A \times A$ has n^2 elements when A has n elements, and a set with m elements has 2^m subsets, there are $2^{|A|^2}$ subsets of $A \times A$. Therefore, there are $2^{|A|^2}$ relations on a set A.

Binary Relations on a Set (cont.)

Example: Consider these relations on the set of integers:

$$R_1 = \overline{\{(a,b) \mid a \le b\}}, \qquad R_4 = \{(a,b) \mid a = b\}, R_2 = \{(a,b) \mid a > b\}, \qquad R_5 = \{(a,b) \mid a = b + 1\}, R_3 = \{(a,b) \mid a = b \text{ or } a = -b\}, \qquad R_6 = \{(a,b) \mid a + b \le 3\}.$$

Note that these relations are on an infinite set and each of these relations is an infinite set.

Which of these relations contain each of the pairs

$$(1,1)$$
, $(1, 2)$, $(2, 1)$, $(1, -1)$, and $(2, 2)$?

Solution: Checking the conditions that define each relation, we see that the pair (1,1) is in R_1 , R_3 , R_4 , and R_6 : (1,2) is in R_1 and R_6 : (2,1) is in R_2 , R_5 , and R_6 : (1,-1) is in R_2 , R_3 , and R_6 : (2,2) is in R_1 , R_3 , and R_4 .

Reflexive Relations

Definition: R is *reflexive* iff $(a,a) \in R$ for every element $a \in A$. Written symbolically, R is reflexive if and only if

$$\forall x[x \in U \longrightarrow (x,x) \in R]$$

Example: The following relations on the integers are reflexive:

$$R_1 = \{(a,b) \mid a \le b\},\$$
 $R_3 = \{(a,b) \mid a = b \text{ or } a = -b\},\$
 $R_4 = \{(a,b) \mid a = b\}.$

If $A = \emptyset$ then the empty relation is reflexive vacuously. That is the empty relation on an empty set is reflexive!

The following relations are not reflexive:

$$R_2 = \{(a,b) \mid a > b\}$$
 (note that $3 \ge 3$),
 $R_5 = \{(a,b) \mid a = b + 1\}$ (note that $3 \ne 3 + 1$),
 $R_6 = \{(a,b) \mid a + b \le 3\}$ (note that $4 + 4 \le 3$).

Symmetric Relations

Definition: R is *symmetric* iff $(b,a) \in R$ whenever $(a,b) \in R$ for all $a,b \in A$. Written symbolically, R is symmetric if and only if

$$\forall x \forall y \ [(x,y) \in R \longrightarrow (y,x) \in R]$$

Example: The following relations on the integers are symmetric:

$$R_3 = \{(a,b) \mid a = b \text{ or } a = -b\},\$$

$$R_4 = \{(a,b) \mid a = b\},\$$

$$R_6 = \{(a,b) \mid a+b \le 3\}.$$

The following are not symmetric:

$$R_1 = \{(a,b) \mid a \le b\}$$
 (note that $3 \le 4$, but $4 \le 3$),

$$R_2 = \{(a,b) \mid a > b\}$$
 (note that $4 > 3$, but $3 \gg 4$),

$$R_5 = \{(a,b) \mid a = b + 1\}$$
 (note that $4 = 3 + 1$, but $3 \neq 4 + 1$).

Antisymmetric Relations

Definition:A relation R on a set A such that for all $a,b \in A$ if $(a,b) \in R$ and $(b,a) \in R$, then a = b is called *antisymmetric*. Written symbolically, R is antisymmetric if and only if

$$\forall x \forall y \ [(x,y) \in R \land (y,x) \in R \longrightarrow x = y]$$

Example: The following relations on the integers are antisymmetric:

$$R_1 = \{(a,b) \mid a \le b\},\$$

$$R_2 = \{(a,b) \mid a > b\},\$$

$$R_4 = \{(a,b) \mid a = b\},\$$

$$R_5 = \{(a,b) \mid a = b + 1\}.$$

The following relations are not antisymmetric:

$$R_3 = \{(a,b) \mid a = b \text{ or } a = -b\}$$
 (note that both (1,-1) and (-1,1) belong to R_3), $R_6 = \{(a,b) \mid a+b \le 3\}$ (note that both (1,2) and (2,1) belong to R_6).

Transitive Relations

Definition: A relation R on a set A is called transitive if whenever $(a,b) \in R$ and $(b,c) \in R$, then $(a,c) \in R$, for all $a,b,c \in A$. Written symbolically, R is transitive if and only if

$$\forall x \forall y \ \forall z [(x,y) \in R \land (y,z) \in R \longrightarrow (x,z) \in R]$$

Example: The following relations on the integers are transitive:

$$R_1 = \{(a,b) \mid a \le b\},\$$
 $R_2 = \{(a,b) \mid a > b\},\$
 $R_3 = \{(a,b) \mid a = b \text{ or } a = -b\},\$
 $R_4 = \{(a,b) \mid a = b\}.$

For every integer, $a \le b$ and $b \le c$, then $b \le c$.

The following are not transitive:

 $R_5 = \{(a,b) \mid a = b + 1\}$ (note that both (3,2) and (4,3) belong to R_5 , but not (4,2)),

 $R_6 = \{(a,b) \mid a+b \le 3\}$ (note that both (2,1) and (1,2) belong to R_6 , but not (2,2)).

Combining Relations

Given two relations R_1 and R_2 , we can combine them using basic set operations to form new relations such as $R_1 \cup R_2$, $R_1 \cap R_2$, $R_1 - R_2$, and $R_2 - R_1$. Example: Let $A = \{1,2,3\}$ and $B = \{1,2,3,4\}$. The relations $R_1 = \{(1,1),(2,2),(3,3)\}$ and $R_2 = \{(1,1),(1,2),(1,3),(1,4)\}$ can be combined using basic set operations to form new relations:

$$R_1 \cup R_2 = \{(1,1),(1,2),(1,3),(1,4),(2,2),(3,3)\}$$

 $R_1 \cap R_2 = \{(1,1)\}$
 $R_1 - R_2 = \{(2,2),(3,3)\}$
 $R_2 - R_1 = \{(1,2),(1,3),(1,4)\}$

Composition

Definition: Suppose

- $-R_1$ is a relation from a set A to a set B.
- $-R_2$ is a relation from B to a set C.

Then the *composition* (or *composite*) of R_2 with R_1 , is a relation from A to C where

-If (x,y) is a member of R_1 and (y,z) is a member of R_2 , then (x,z) is a member of $R_2 \circ R_1$.

```
Example: What is the composite relation of R and S?
```

R:
$$\{1,2,3\} \rightarrow \{1,2,3,4\}$$
 S: $\{1,2,3,4\} \rightarrow \{0,1,2\}$ R = $\{(1,1),(1,4),(2,3),(3,1),(3,4)\}$ S = $\{(1,0),(2,0),(3,1),(3,2),(4,1)\}$

S o R =
$$\{(1,0), (3,0), (1,1), (3,1), (2,1), (2,2)\}$$

Powers of a Relation

Definition: Let R be a binary relation on A. Then the powers R^n of the relation R can be defined inductively by:

- Basis Step: $R^1 = R$
- − Inductive Step: $R^{n+1} = R^n \circ R$

The powers of a transitive relation are subsets of the relation. This is established by the following theorem:

Theorem 1: The relation R on a set A is transitive iff $R^n \subseteq R$ for $n = 1,2,3 \ldots$

(see the text for a proof via mathematical induction)



N-ary Relations

Definition: Let A_1 , A_2 , ..., A_n be sets. An n-ary relation on these sets is a subset of $A_1 \times A_2 \times ... \times A_n$. The sets A_1 , A_2 , ..., A_n are called the domains of the relation, and n is its degree.

Data is stored in relations (a.k.a., tables)

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

Enrollment		
Stud_ID	Course	
334322	CS 441	
334322	Math 336	
546346	Math 422	
964389	Art 707	

Columns of a table represent the attributes of a relation

Rows, or records, contain the actual data defining the relation



Selection operator – Filter rows in a table

Definition: Let R be an n-ary relation and let C be a condition that elements in R must satisfy. The selection s_c maps the n-ary relation R to the n-ary relation of all n-tuples from R that satisfy the condition C.

Example: Consider the Students relation from earlier in lecture. Let the condition C1 be Major="CS" and let C2 be GPA > 2.5. What is the result of $s_{C1 \land C2}$ (Students)?

Answer:

- (Alice, 334322, CS, 3.45)
- (Charlie, 045628, CS, 2.75)

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

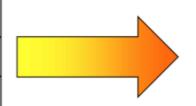
Projection operator – allows to consider only the subset of columns of table



Definition: The projection $P_{i1,...,in}$ maps the n-tuple $(a_1, a_2, ..., a_n)$ to the m-tuple $(a_{i1}, ..., a_{im})$ where $m \le n$

Example: What is the result of applying the projection $P_{1,3}$ to the Students table?

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0



Name	Major
Alice	CS
Bob	Math
Charlie	CS
Denise	Art

Join or Equijoin operator – Create a new table based on data from two or more related tables



Definition: Let R be a relation of degree m and S be a relation of degree n. The equijoin $J_{i1=j1,...,ik=jk}$, where $k \le m$ and $k \le n$, creates a new relation of degree m+n-k containing the subset of S × R in which $s_{i1} = r_{j1}$, ..., $s_{ik} = r_{jk}$ and duplicate columns are removed (via projection).

Example: What is the result of the equijoin $J_{2=1}$ on the Students and Enrollment tables?

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

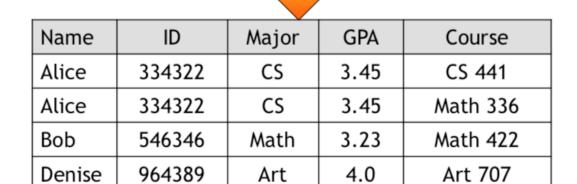
Enrollment		
Stud_ID	Course	
334322	CS 441	
334322	Math 336	
546346	Math 422	
964389	Art 707	



Join/Equijoin operator

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0

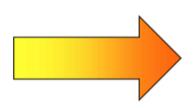
Enrollment		
Stud_ID	Course	
334322	CS 441	
334322	Math 336	
546346	Math 422	
964389	Art 707	





SQL (Structured Query Language)

Students			
Name	ID	Major	GPA
Alice	334322	CS	3.45
Bob	546346	Math	3.23
Charlie	045628	CS	2.75
Denise	964389	Art	4.0



Name	ID
Alice	334322
Charlie	045628

SELECT Name, ID FROM Students WHERE Major = "CS" AND

GPA > 2.5↑

SELECT is actually a projection (in this case, P_{1,2})

The WHERE clause lets us filter (i.e., $S_{major="CS" \land GPA>2.5}$)

Representing Relations Using Matrices

A relation between finite sets can be represented using a zero-one matrix.

Suppose *R* is a relation from $A = \{a_1, a_2, ..., a_m\}$ to $B = \{b_1, b_2, ..., b_n\}$.

- The elements of the two sets can be listed in any particular arbitrary order. When A = B, we use the same ordering.
- The relation R is represented by the matrix $M_R = [m_{ij}]$, where

$$m_{ij} = \begin{cases} 1 \text{ if } (a_i, b_j) \in R, \\ 0 \text{ if } (a_i, b_j) \notin R. \end{cases}$$

The matrix representing R has a 1 as its (i,j) entry when a_i is related to b_j and a 0 if a_i is not related to b_j .

Examples of Representing Relations Using innovated Matrices

Example 1: Suppose that $A = \{1,2,3\}$ and $B = \{1,2\}$. Let R be the relation from A to B containing (a,b) if $a \in A$, $b \in B$, and a > b. What is the matrix representing R (assuming the ordering of elements is the same as the increasing numerical order)?

Solution: Because $R = \{(2,1), (3,1), (3,2)\}$, the matrix is

$$M_R = \left[egin{array}{ccc} 0 & 0 \ 1 & 0 \ 1 & 1 \end{array}
ight].$$

Examples of Representing Relations Using Matrices

Example 2: Let $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2, b_3, b_4, b_5\}$. Which ordered pairs are in the relation R represented by the matrix

$$M_R = \left[egin{array}{ccccc} 0 & 1 & 0 & 0 & 0 \ 1 & 0 & 1 & 1 & 0 \ 1 & 0 & 1 & 0 & 1 \end{array}
ight]?$$

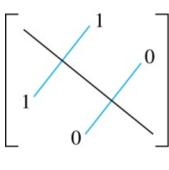
Solution: Because R consists of those ordered pairs (a_i,b_j) with $m_{ij} = 1$, it follows that:

$$R = \{(a_1, b_2), (a_2, b_1), (a_2, b_3), (a_2, b_4), (a_3, b_1), \{(a_3, b_3), (a_3, b_5)\}.$$

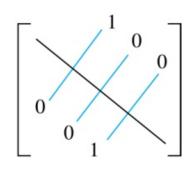
Matrices of Relations on Sets

If R is a reflexive relation, all the elements on the main diagonal of M_R are equal to 1.

R is a symmetric relation, if and only if $m_{ij} = 1$ whenever $m_{ji} = 1$. R is an antisymmetric relation, if and only if $m_{ij} = 0$ or $m_{ji} = 0$ when $i \neq j$.







(b) Antisymmetric

Example of a Relation on a Set

Example 3: Suppose that the relation *R* on a set is represented by the matrix

$$M_R = \left[egin{array}{cccc} 1 & 1 & 0 \ 1 & 1 & 1 \ 0 & 1 & 1 \end{array}
ight].$$

Is *R* reflexive, symmetric, and/or antisymmetric?

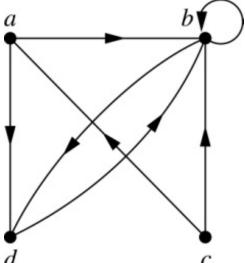
Solution: Because all the diagonal elements are equal to 1, R is reflexive. Because M_R is symmetric, R is symmetric and not antisymmetric because both $m_{1,2}$ and $m_{2,1}$ are 1.

Representing Relations Using Digraphs

Definition: A *directed graph*, or *digraph*, consists of a set V of *vertices* (or *nodes*) together with a set E of ordered pairs of elements of V called *edges* (or *arcs*). The vertex a is called the *initial vertex* of the edge (a,b), and the vertex b is called the *terminal vertex* of this edge.

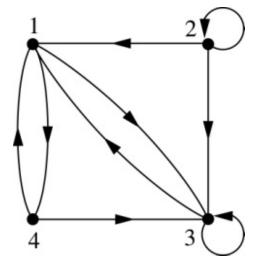
– An edge of the form (*a*,*a*) is called a *loop*.

Example 7: A drawing of the directed graph with vertices a, b, c, and d, and edges (a, b), (a, d), (b, b), (b, d), (c, a), (c, b), and (d, b) is shown here.



Examples of Digraphs Representing Relations

Example 8: What are the ordered pairs in the relation represented by this directed graph?



Solution: The ordered pairs in the relation are

Properties

Reflexivity: A loop must be present at all vertices in the graph.

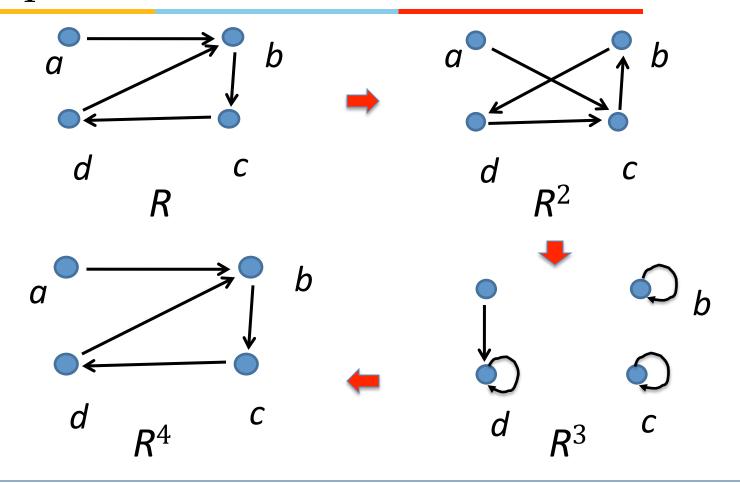
Symmetry: If (x,y) is an edge, then so is (y,x).

Antisymmetry: If (x,y) with $x \neq y$ is an edge, then (y,x) is not an edge.

Transitivity: If (x,y) and (y,z) are edges, then so is (x,z).

lead

Example of the Powers of a Relation



The pair (x,y) is in \mathbb{R}^n if there is a path of length n from x to y in \mathbb{R} (following the direction of the arrows).

Reflexive closure

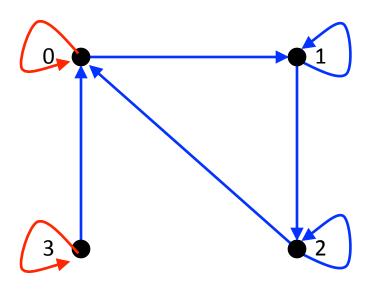
In order to find the reflexive closure of a relation *R*, we add a loop at each node that does not have one

The reflexive closure of R is R U Δ

- -Where $\Delta = \{ (a,a) \mid a \in R \}$
 - Called the "diagonal relation"
- -With matrices, we set the diagonal to all 1's

Reflexive closure

Let R be a relation on the set $\{0, 1, 2, 3\}$ containing the ordered pairs (0,1), (1,1), (1,2), (2,0), (2,2), and (3,0) What is the reflexive closure of R? We add all pairs of edges (a,a) that do not already exist



We add edges: (0,0), (3,3)



Symmetric closure

In order to find the symmetric closure of a relation R, we

add an edge from a to b, where there is already an edge

from b to a

The symmetric closure of R is $R \cup R^{-1}$

If
$$R = \{ (a,b) \mid ... \}$$

Then $R^{-1} = \{ (b,a) \mid ... \}$

Let R be a relation on the set $\{0, 1, 2, 3\}$ containing the ordered pairs (0,1), (1,1), (1,2), (2,0), (2,2), and (3,0)

What is the symmetric closure of *R*?

We add all pairs of edges (a,b) where (b,a) exists

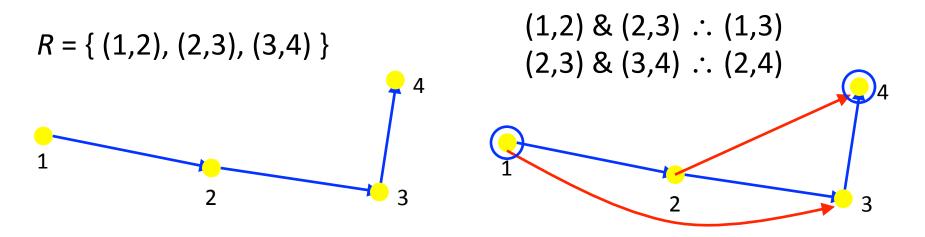
We make all "single" edges into anti-parallel pairs

We add edges: (0,2), (0,3) (1,0), (2,1)

Transitive closure

Informal definition: If there is a path from *a* to *b*, then there should be an edge from *a* to *b* in the transitive closure First take of a definition:

In order to find the transitive closure of a relation R, we add an edge from a to c, when there are edges from a to b and b to c But there is a path from 1 to 4 with no edge!





Transitive Closure – Algorithm 1

Let R be a relation on a set A. The *connectivity relation* R^* consists of the pairs (a, b) such that there is a path of length at least one from a to b in R.

$$R^* = \bigcup_{n=1}^{\infty} R^n.$$

THEOREM 2

The transitive closure of a relation R equals the connectivity relation R^* .

Compute $\mathbf{M}_{R}^{[2]} = \mathbf{M}_{R} \circ \mathbf{M}_{R}$ Join that with \mathbf{M}_{R} to yield $\mathbf{M}_{R} \vee \mathbf{M}_{R}^{[2]}$ Compute $\mathbf{M}_{R}^{[3]} = \mathbf{M}_{R}^{[2]} \circ \mathbf{M}_{R}$ Join that with $\mathbf{M}_{R} \vee \mathbf{M}_{R}^{[2]}$ from above

$$\mathbf{M}_{R^*} = \mathbf{M}_R \vee \mathbf{M}_R^{[2]} \vee \mathbf{M}_R^{[3]} \vee \cdots \vee \mathbf{M}_R^{[n]}$$

lead

$$M_{R} = \begin{bmatrix} 4 & 6 & 8 & 10 \\ 4 & 1 & 0 & 0 & 1 \\ 6 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Fig. 16.

Now find the powers of M_R as in Figs. 17, 18 and 19.

$$\mathbf{M}_{R^{2}} = \begin{matrix} 4 & 6 & 8 & 10 \\ 4 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 10 & 0 & 0 & 0 \end{matrix} \\ \mathbf{Fig. 17.} \qquad \qquad \begin{matrix} 4 & 6 & 8 & 10 \\ 4 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} \\ \mathbf{Fig. 18.}$$

$$\mathbf{M_{R^4}} = \begin{bmatrix} 4 & 6 & 8 & 10 \\ 4 & 1 & 0 & 0 & 1 \\ 6 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 10 & 0 & 0 & 0 & 0 \end{bmatrix}$$

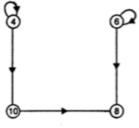
Hence, the transitive closure of M_R is M_{R^*} as shown in Fig. 20. (where M_{R^*} is the ORing of powers of

$$M_{R'} = M_{R} \vee M_{R^{2}} \vee M_{R^{3}} \vee M_{R^{4}}; \qquad M_{R'} = \begin{bmatrix} 4 & 6 & 8 & 10 \\ 4 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 10 & 0 & 0 & 0 \end{bmatrix}$$
Fig. 20.

Thus, $R^* = \{(4, 4), (4, 10), (6, 8), (6, 6), (6, 10), (8, 10)\}.$

 M_R).

lead



$$\mathbf{M}_{R} = \begin{bmatrix} 4 & 6 & 8 & 10 \\ 4 & 1 & 0 & 0 & 1 \\ 6 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 10 & 0 & 0 & 0 \end{bmatrix}$$

Fig. 21.

The value of n = 1 A 1 = 4. Thus, we have to find the Warshall's sets w_0 , w_1 , w_2 , w_3 and w_4 . The first set w_0 is same as M_R , which is shown below:

$$w_0 = M_R = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Now to find w_1 from w_0 we have row number 1 for column 1 in w_0 and column number 1 and 4 for row 1 in w_0 . Thus, new entries in w_1 are (4, 4) and (4, 10) which are already one. Thus, w_1 is same as w_0 , which is as follows:

$$w_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

To find w_2 from w_1 , we have row number 2 for column 2 in w_1 and column numbers 2 and 3 for row 2 in w_1 . Thus, new entries in w_2 are (6, 6) and (6, 8), which are already one. Thus, w_2 is same as w_1 , which is as follows:

$$w_2 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Similarly, w_3 is obtained from w_2 . Here, we have row number 2 in column 3 and column number 4 in row 3. Thus, the new entries in w_3 are (6, 10). So w_3 is as follows:

Warshall's Algorithm

$$w_3 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Similarly, w_4 is obtained from w_3 . But there are no new entries of 1's in w_4 . Hence $M_R = w_4 = w_3$,

Operation Count

Algorithm 1: To find the transitive closure, Algorithm 1 uses $2n^3(n-1)$ bit operations

Warshall's Algorithm:

To find the transitive closure, Warshall's algorithm uses 2n³ bit operations

Warshall's Algorithm

ALGORITHM 2 Warshall Algorithm.

```
\begin{aligned} \mathbf{procedure} \ \textit{Warshall} \ (\mathbf{M}_R : n \times n \ \text{zero--one matrix}) \\ \mathbf{W} := \mathbf{M}_R \\ \mathbf{for} \ k := 1 \ \mathbf{to} \ n \\ \mathbf{for} \ i := 1 \ \mathbf{to} \ n \\ \mathbf{for} \ i := 1 \ \mathbf{to} \ n \\ w_{ij} := u_{ij} \vee (w_{ik} \wedge w_{kj}) \\ \mathbf{return} \ \mathbf{W} \{\mathbf{W} = [w_{ij}] \ \text{is} \ \mathbf{M}_{R^*} \} \end{aligned}
```