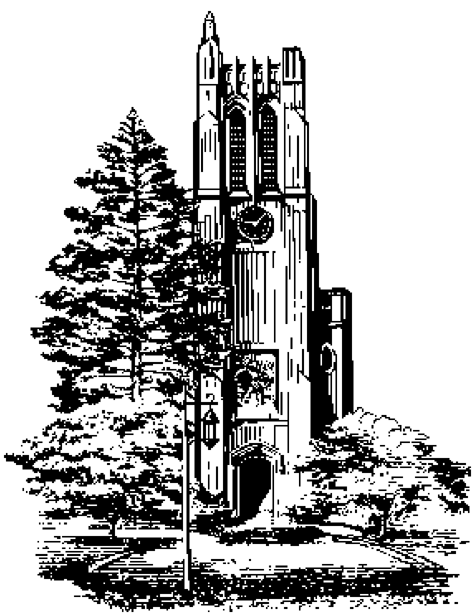
Closures of a Relation



N Geetha
AM & CS
PSG Tech

Closures of Relations

- "Closing" a relation has important applications.
- The natural numbers ***N*** (counting numbers) are not closed under subtraction: when we close them under subtraction, we get ***Z***, the integers (positive and negative).
- When we close ***Z*** under the operation of division, we get ***Q***, the rational numbers.

Applications of Closures

- Computing the transitive closure of a digraph is an important problem in many computer science applications:
  - Evaluation of recursive database queries.
  - Analysis of reachability (connectivity) of transition graphs in communication networks.
  - Construction of parsing automata in compilers.

Introduction

- *Closure* of relation properties
  - When a property does not hold for a relation, how could we *minimally* augment the relation so that the property would hold?

Reflexive Closure

- Example: Consider the relation
  R = {(1,1), (1,2), (2,1), (3,2)} on set {1,2,3}
  - Is it reflexive?
  - How can we produce a reflexive relation containing R that is as small as possible?

Reflexive Closure – cont.

- When a relation R on a set A is not reflexive:
  - How to minimally augment R (adding the minimum number of ordered pairs) to make it a reflexive relation?
    - The reflexive closure of R.
- The reflexive closure of R can be formed by adding all of the pairs of the form (a,a) to R. In other words we should find:

$$R \cup \Delta = R \cup \{(a,a) \mid a \in A\}$$

R = { (1,1), (1,2), (2,1), (3,2) } U { (1,1), (2,2), (3,3) }

Reflexive Closure – Cont.

- The diagonal relation on A is:

$$\Delta = \{(a,a) \mid a \in A\}.$$

- The reflexive closure of R is then:

$$R \cup \Delta.$$

- <u>Properties:</u>
  - $R \subseteq (R \cup \Delta)$;
  - $R \cup \Delta$ is reflexive;
  - $\forall S \quad (R \subseteq S \ \wedge \ S \text{ is reflexive}) \ \rightarrow \ (R \cup \Delta) \subseteq S$.
- In zero-one matrix notation: $M_R \vee M_\Delta$
- Turn on all the diagonal bits!

Example 1

- Consider $R = \{(a,b) \in \mathbf{Z}\times\mathbf{Z} \mid a < b\}$
- The reflexive closure of relation R is:

$R \cup \Delta$

$= \{(a,b) \in \mathbf{Z}\times\mathbf{Z} \mid a < b\} \cup \{(a,a) \mid a \in \mathbf{Z}\}$

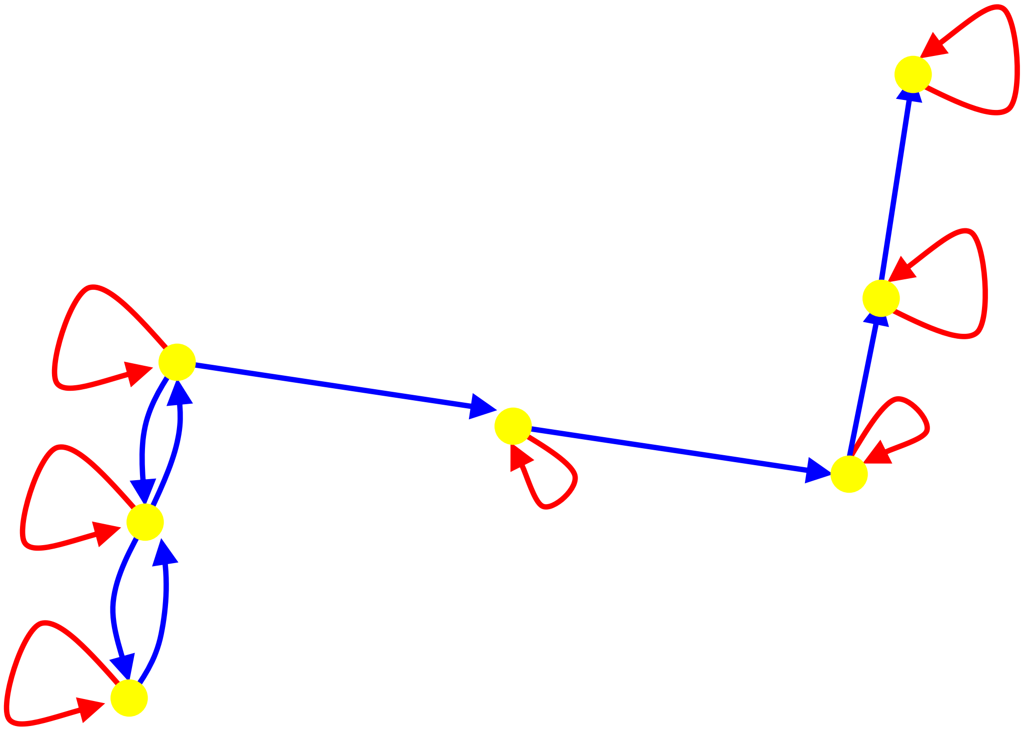
$= \{(a,b) \in \mathbf{Z}\times\mathbf{Z} \mid a \leq b\}$

Reflexive closure

- Consider a relation *R*:
  - Note that it is not reflexive

- We want to add edges to make the relation reflexive

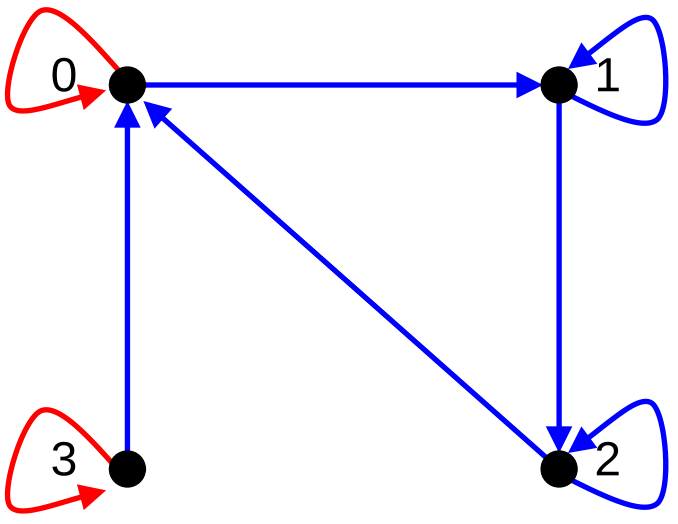
- By adding those edges, we have made a non-reflexive relation *R* into a reflexive relation



- This new relation is called the **reflexive closure** of *R*

Reflexive closure Example 2

- 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

- Example: Consider
  R ={(1,1), (1,2), (2,2), (2,3), (3,1), (3,2)}
  - R is not symmetric; the pairs missing are: (2,1), (1,3).
  - If we add those, we obtain the new relation:
    {(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2)}.
    The new relation is symmetric.

Symmetric Closure

- When a relation *R* on a set *A* is not symmetric:
  – How to minimally augment *R* (adding the minimum number of ordered pairs) to have a symmetric relation?
  – The symmetric closure of *R*.

Symmetric Closure

- The inverse of R is:

$$R^{-1} = \{(b,a) : (a, b) \in R\}.$$

- The symmetric closure of R is then:

$$R \cup R^{-1}.$$

- Properties:
  - $R \subseteq (R \cup R^{-1})$;
  - $R \cup R^{-1}$ is symmetric;
  - $\forall S \quad (R \subseteq S \;\wedge\; S \text{ is symmetric}) \;\rightarrow\; (R \cup R^{-1}) \subseteq S$.

- In zero-one matrix notation: $M_R \vee M_R^{t}$

Example 1

- Consider $R = \{(a,b) \in \mathbf{Z} \times \mathbf{Z} \mid a < b\}$
- The symmetric closure of relation R is:

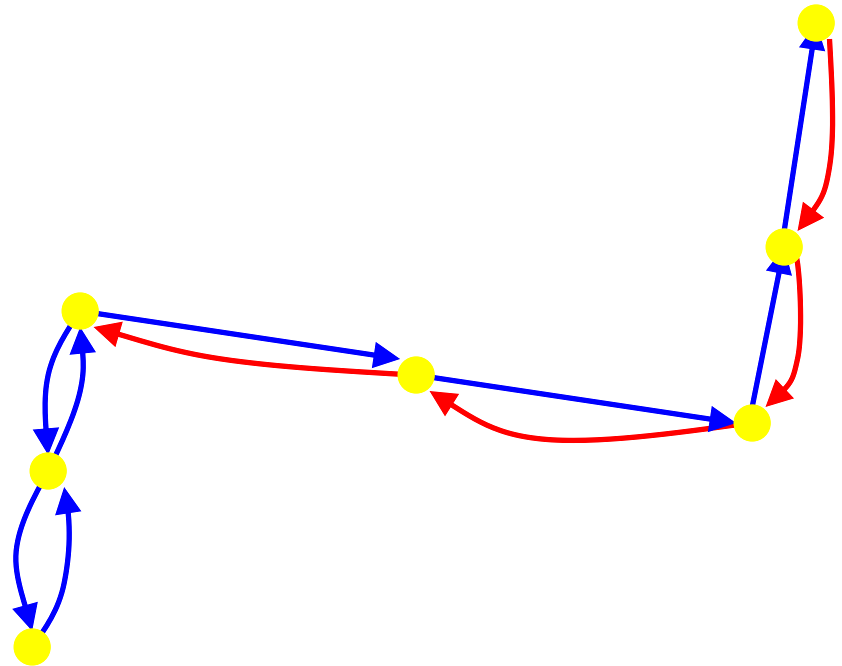
$$R \cup R^{-1}$$

$$= \{(a,b) \in \mathbf{Z} \times \mathbf{Z} \mid a < b\} \cup \{(a,b) \in \mathbf{Z} \times \mathbf{Z} \mid b < a\}$$

$$= \{(a,b) \in \mathbf{Z} \times \mathbf{Z} \mid a \neq b\}$$

Symmetric closure

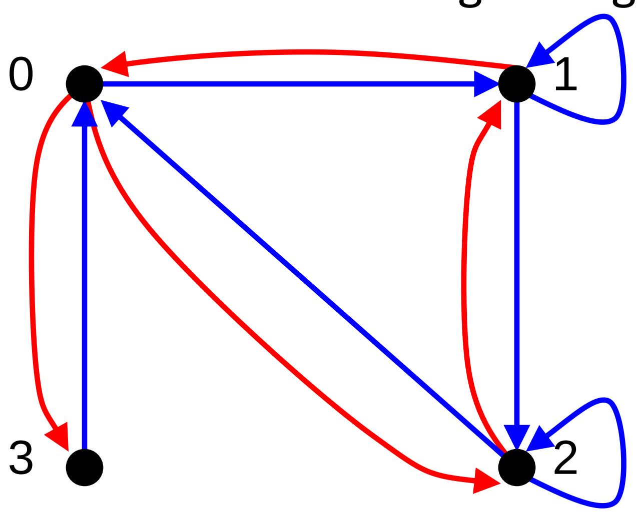
- Consider a relation *R*:
  - Note that it is not symmetric
- We want to add edges to make the relation symmetric
- By adding those edges, we have made a non-symmetric relation *R* into a symmetric relation



- This new relation is called the **symmetric closure** of *R*

Example

- 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

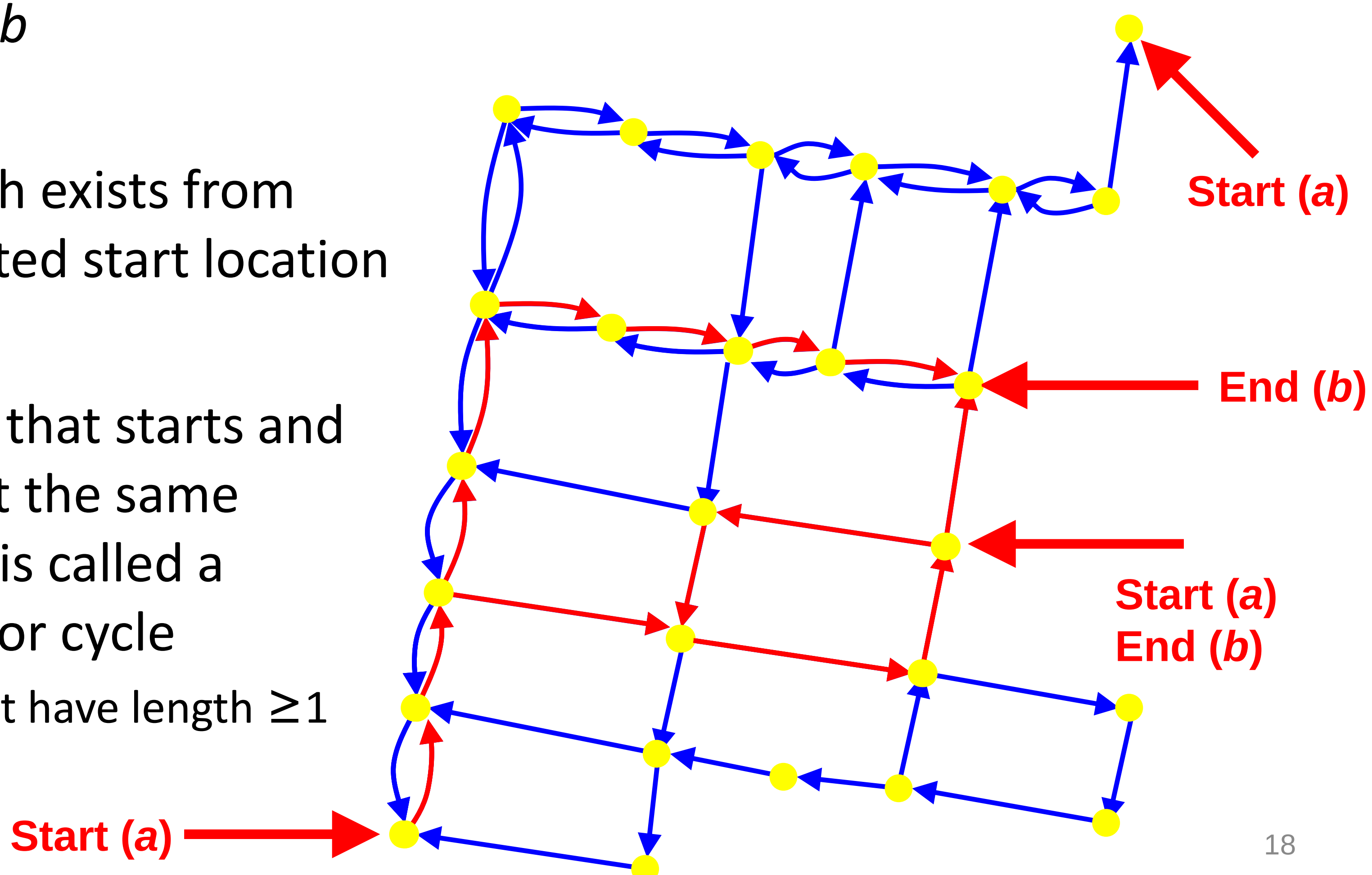
- Consider $R = \{(1,3), (1,4), (2,1), (3,2)\}$.
  - R is not transitive;
  - What are the missing terms?
    - Few are: (1,2), (2,3), (2,4), (3,1).
  - If we add those, we obtain the new relation:

    {(1,2), (1,3), (1,4), (2,1), (2,3), (2,4), (3,1), (3,2)}.
  - Is the above relation transitive?
  - No, it is not. Missing terms such as (1,1), (2,2)…

➢ Transitive closure is more complicated to build.

Paths in directed graphs

- A *path* is a sequences of connected edges from vertex *a* to vertex *b*
- No path exists from the noted start location
- A path that starts and ends at the same vertex is called a circuit or cycle
  - Must have length ≥1



Paths in Directed Graphs

- The length of a path is the number of **edges** in the path, not the number of nodes
- Definition: A *path* from a to b in a digraph G is a sequence of 1 or more adjacent arcs

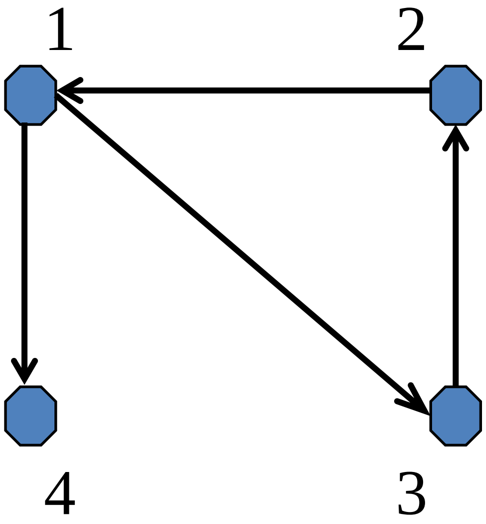
$$(a,x_1), (x_1,x_2), (x_2,x_3), \ldots, (x_{n-1}, b).$$

  – Denoted: $a, x_1, x_2, x_3, \ldots, x_{n-1}, b$
  – Has *length* n.
  – If $a = b$, the path is called a *circuit* or a *cycle,* since the path returns to its start.

Does a path appear to be “transitive”?

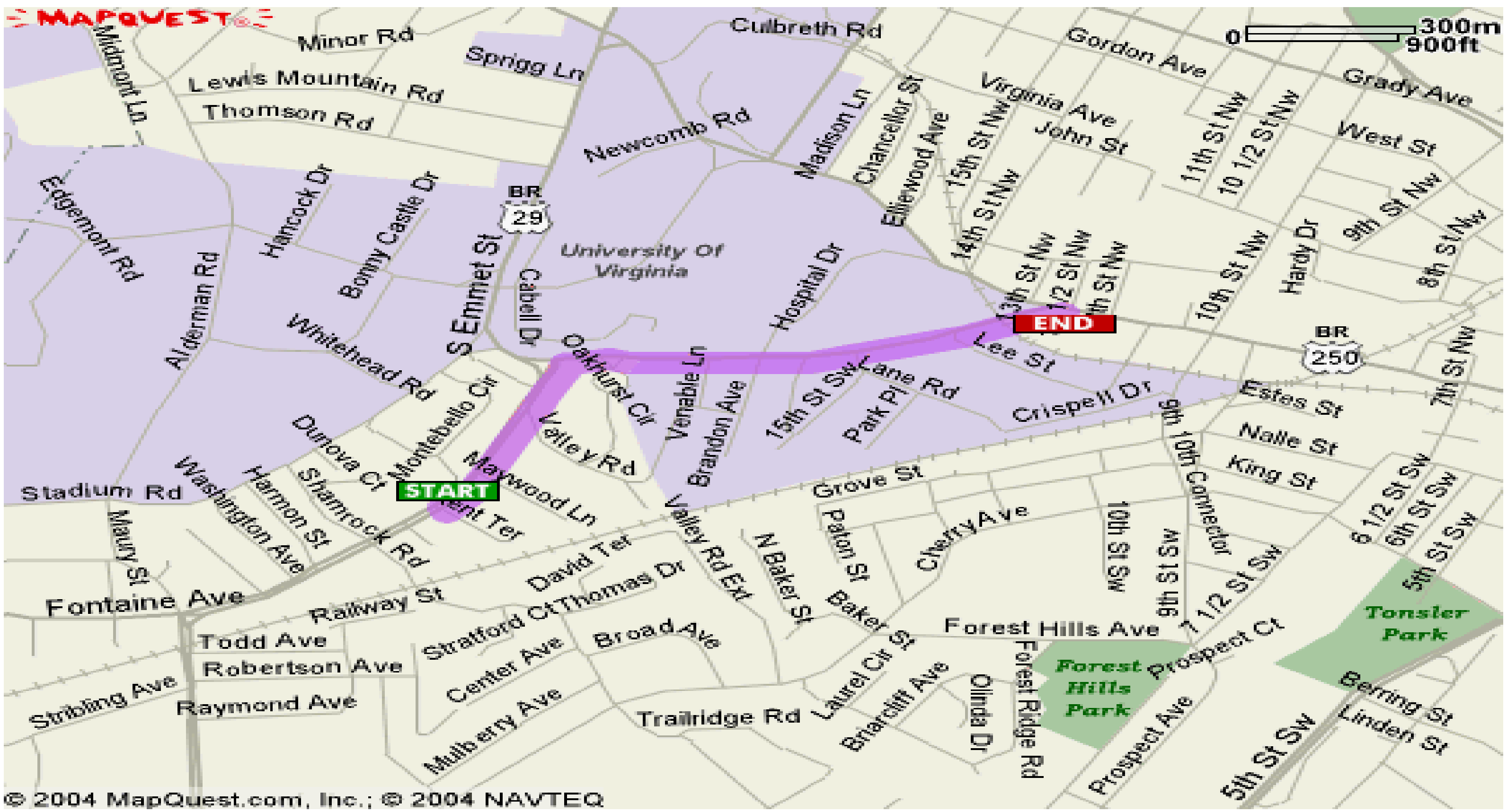
Paths in Relations

- There is a path from a to b in R if there is a sequence of elements: $x_1, x_2, \ldots, x_{n-1}$ with $(a,x_1)\in R$, $(x_1,x_2)\in R$, …, $(x_{n-1}, b)\in R$.
- Example. $R = \{(1,3), (1,4), (2,1), (3,2)\}$
  - $\exists$ a path in R (len. 2) from 1 to 2: $(1,3), (3,2)\in R$.
  - $\exists$ a path in R (len. 3) from 3 to 4: $(3,2), (2,1), (1,4)\in R$.
  - There is no path in R from 4 to 1, nor to 2, nor to 3.



Shortest paths

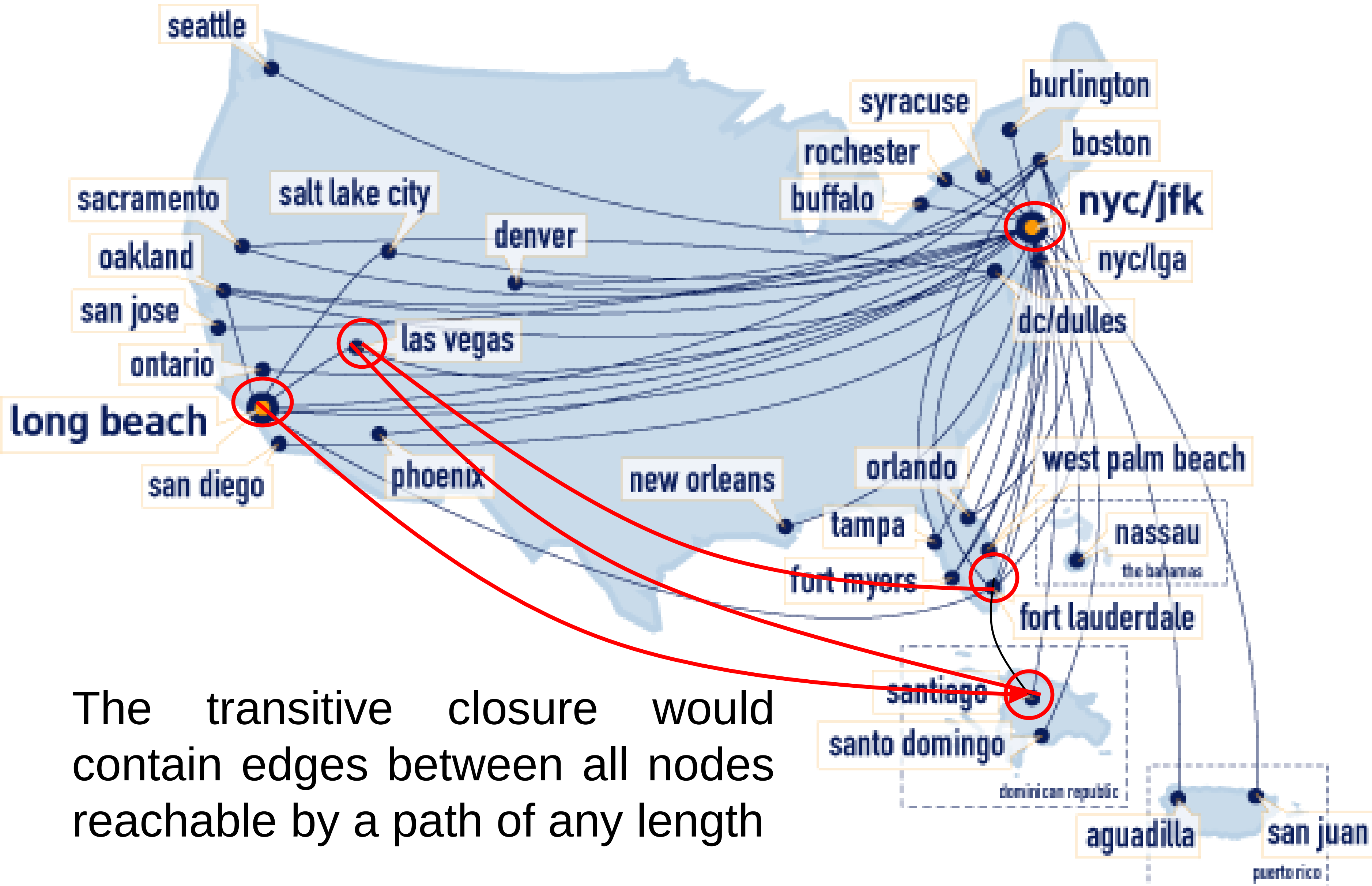
- What is really needed in most applications is finding the shortest path between two vertices



Transitive Closure

- When a relation R on a set A is not transitive:
  - How to minimally augment R (adding the minimum number of ordered pairs) to have a transitive relation?
  - The transitive closure of R.

Transitive closure

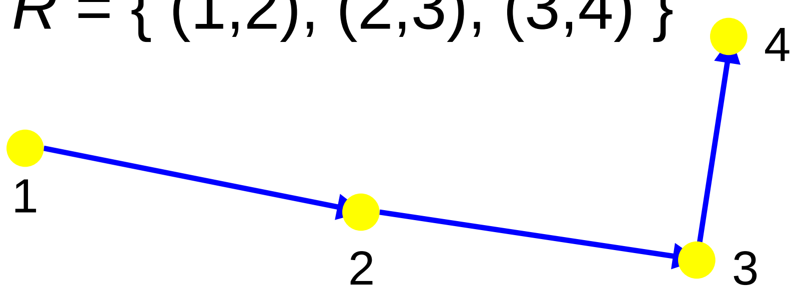


The transitive closure would contain edges between all nodes reachable by a path of any length

Transitive closure

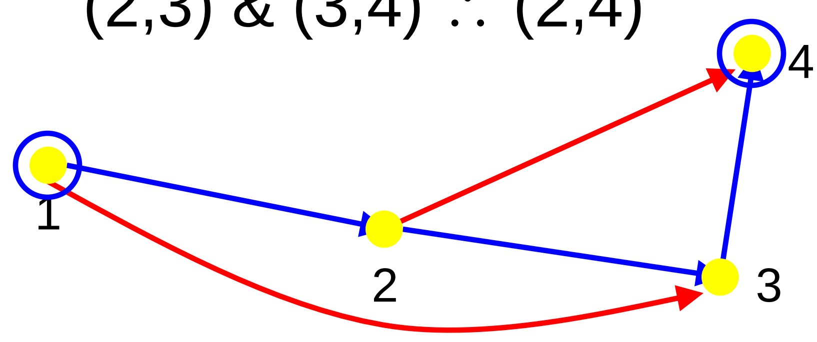
- Informal definition: If there is a path from *a* to *b*, then there should be an edge from *a* to *b* in the transitive closure
- First take of a definition:
  - In order to find the transitive closure of a relation *R*, we add an edge from *a* to *c*, when there are edges from a to *b* and *b* to *c*
- But there is a path from 1 to 4 with no edge!

$R = \{ (1,2), (2,3), (3,4) \}$



$(1,2)$ & $(2,3)$ $\therefore$ $(1,3)$

$(2,3)$ & $(3,4)$ $\therefore$ $(2,4)$

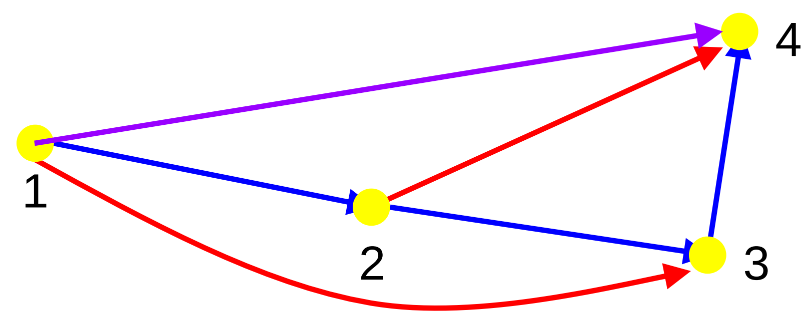


Transitive closure

- 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

- Second 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*
  - Repeat this step until no new edges are added to the relation
- We will study different algorithms for determining the transitive closure

- red means added on the first repeat
- purple means added on the second repeat



6 degrees of separation

- The idea that everybody in the world is connected by six degrees of separation
  - Where 1 degree of separation means you know (or have met) somebody else
- Let R be a relation on the set of all people in the world
  - $(a,b) \in R$ if person a has met person b
- So six degrees of separation for *any* two people a and g means:
  - (a,b), (b,c), (c,d), (d,e), (e,f), (f,g) are all in R
- Or, $(a,g) \in R^6$

Connectivity relation

- R contains edges between all the nodes reachable via 1 edge
- $R \circ R = R^2$ contains edges between nodes that are reachable via 2 edges in R
- $R^2 \circ R = R^3$ contains edges between nodes that are reachable via 3 edges in R
- R^n = contains edges between nodes that are reachable via n edges in R

- R^* contains edges between nodes that are reachable via any number of edges (i.e. via any path) in R
  - Rephrased: R^* contains all the edges between nodes a and b when there is a path of length at least 1 between a and b in R

- R^* is the transitive closure of R
  - The definition of a transitive closure is that there are edges between any nodes (a,b) that contain a path between them

Theorem 1

- Theorem: Let R be a relation on a set A. $\exists$ a path in R of length n from a to $b \Leftrightarrow \underline{(a,b) \in R^n}$.

Proof : By Defn., there is a path from a to b of length 1 iff (a,b) ε R. True for n=1.

Assume the theorem is true for n. This is the inductive hypothesis. Prove using MI

Transitive Closure

- Definition: Let R be a relation on a set A.

  The *connectivity relation* is the relation R* defined as:

  $$R^* = \{(a,b) \mid \exists \text{ a path in } R \text{ from } a \text{ to } b\}.$$

  - From the above definition of R^* and the previous theorem, R^* can be written as:

  $$R^* = \bigcup^{\infty}_{k=1} R^k.$$

Transitive Closure – Cont.

- **<u>Theorem 2</u>** The transitive closure of R is the connectivity relation R^*.
- <u>Proof:</u>
  1. $R \subseteq R^*$, because if there is an edge from a to b then there is a path from a to b.
  2. R^* is transitive, because if $\exists$ a path from a to b and $\exists$ a path from b to c then $\exists$ a path from a to c.
  3. $\forall S\ (R \subseteq S \wedge S$ is transitive$) \rightarrow R^* \subseteq S$.

     Proof. Let S be a transitive relation, and assume $R \subseteq S$.

     So $R^* \subseteq S^*$ because $R \subseteq S$. *(Any path in R is also a path in S)*

     Also $\forall k\ S^k \subseteq S$, because S is transitive

     So $S^* \subseteq S$. Therefore, $R^* \subseteq S$.

How long are the paths in a transitive closure?

- Let R be a relation on set A, and let A be a set with n elements
  - Rephrased: consider a graph G with n nodes and some number of edges
- **Lemma 1:** If there is a path (of length at least 1) from a to b in R, then there is a path between a and b of length not exceeding n
- Proof preparation:
  - Suppose there is a path from a to b in R
  - Let the length of that path be m
  - Let the path be edges $(x_0, x_1), (x_1, x_2), \ldots, (x_{m-1}, x_m)$
  - That's nodes $x_0, x_1, x_2, \ldots, x_{m-1}, x_m$
  - If a node exists twice in our path, then it's not a shortest path
    - As we made no progress in our path between the two occurrences of the repeated node
  - Thus, each node may exist at most once in the path

How long are the paths in a transitive closure?

- Proof by contradiction:
  - Assume there are more than n nodes in the path
    - Thus, $m > n$
    - Let $m = n+1$
  - By the pigeonhole principle, there are $n+1$ nodes in the path (pigeons) and they have to fit into the n nodes in the graph (pigeonholes)
  - Thus, there must be at least one pigeonhole that has at least two pigeons
  - Rephrased: there must be at least one node in the graph that has two occurrences in the nodes of the path
    - Not possible, as the path would not be the shortest path
  - Thus, it cannot be the case that $m > n$
- If there exists a path from a to b, then there is a path from a to b of at most length n

Transitive Closure – Cont.

- Lemma: Let A be a set with n elements, and R be a relation on A.

  If there is a path in R from a to b (with $a \neq b$), then there is such a path with length $\leq n$-1.

➢ $R^* = R \cup R^2 \cup R^3 \cup \ldots \cup R^n$.

Transitive Closure – Method 2

Theorem: Let M_R be the zero-one matrix of relation R on a set with n elements.

The zero-one matrix of the transitive closure R^* is:

$$M_{R^*} = M_R \vee M_R^{[2]} \vee M_R^{[3]} \vee \ldots \vee M_R^{[n]}.$$

Finding the transitive closure

- Let $\mathbf{M}_R$ be the zero-one matrix of the relation R on a set with n elements. Then the zero-one matrix of the transitive closure R^* is:

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

Nodes reachable with one application of the relation

Nodes reachable with two applications of the relation

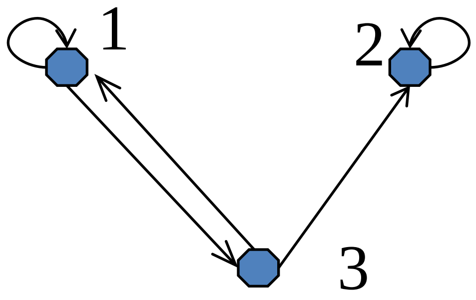
Nodes reachable with *n* applications of the relation

Transitive Closure - Example

- Find the matrix $\mathbf{M}_R$* of the transitive closure of R:

$\mathbf{M}_R$ =

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

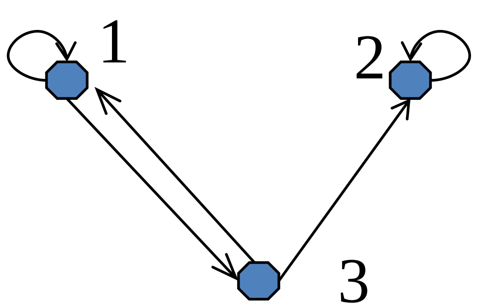


Transitive Closure - Example

- Find the matrix $\mathbf{M}_R^*$ of the transitive closure of R:

$\mathbf{M}_R =$

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



- Solution:

$$\mathbf{M}_R^{[2]} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{M}_R^{[3]} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

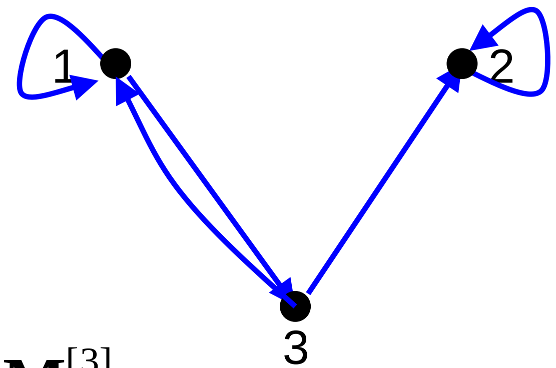
$$\mathbf{M}_R^* = \mathbf{M}_R \vee \mathbf{M}_R^{[2]} \vee \mathbf{M}_R^{[3]} =$$

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

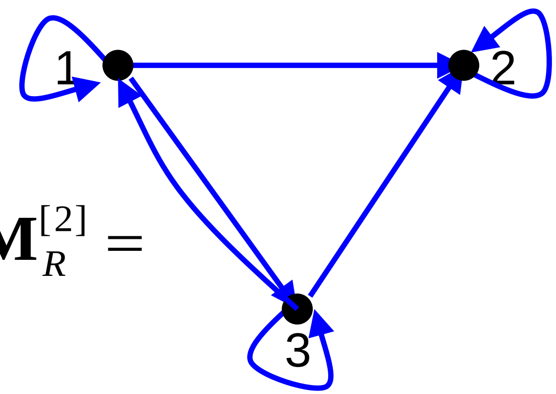
Example in terms of the Digraph

- Find the zero-one matrix of the transitive closure of the relation R given by:

$$\mathbf{M}_R = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$



$$\mathbf{M}_R^{[2]} =$$

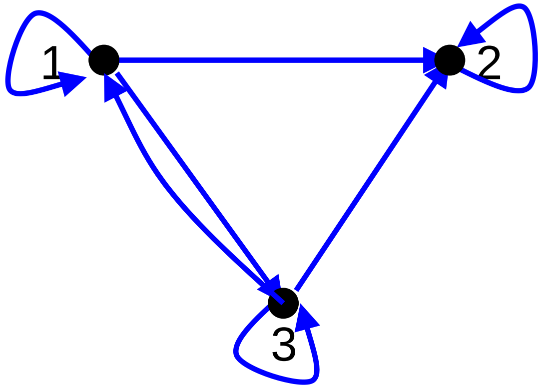


$$\mathbf{M}_{R^*} = \mathbf{M}_R \vee \mathbf{M}_R^{[2]} \vee \mathbf{M}_R^{[3]}$$

$$\mathbf{M}_R^{[2]} = \mathbf{M}_R \odot \mathbf{M}_R = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \odot \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Example in terms of digraph

$$\mathbf{M}_R^{[3]} = \mathbf{M}_R^{[2]} \odot \mathbf{M}_R = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \odot \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$



$$\mathbf{M}_{R^*} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \vee \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \vee \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Transitive closure algorithm 1(Matrix Method)

- What we did :
  - Compute the next matrix $\mathbf{M}_R^{[i]}$, where $1 \le i \le n$
  - Do a Boolean join with the previously computed matrix
- For our example:
  - 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

Transitive closure algorithm

procedure *transitive_closure* ($\mathbf{M}_R$: zero-one $n \times n$ matrix)

  $\mathbf{A} := \mathbf{M}_R$

  $\mathbf{B} := \mathbf{A}$

  for $i := 2$ **to** n

  begin

    $\mathbf{A} := \mathbf{A} \odot \mathbf{M}_R$

    $\mathbf{B} := \mathbf{B} \vee \mathbf{A}$

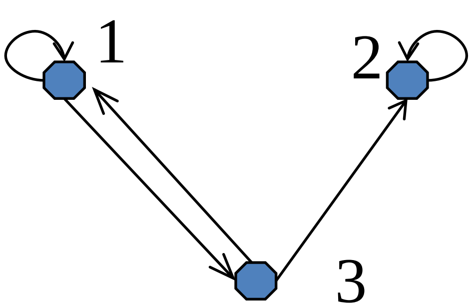
  end { $\mathbf{B}$ is the zero-one matrix for R^* }

Transitive Closure - Example

- Find the matrix $\mathbf{M}_R$* of the transitive closure of R:

$\mathbf{M}_R =$

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



<u>Solution:</u>

$\mathbf{B}_1 = \mathbf{A}_1 = \mathbf{M}_R = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$

$\mathbf{A}_k = \mathbf{A}_{k-1} \circ \mathbf{M}_R \qquad \mathbf{B}_k = \mathbf{B}_{k-1} \vee \mathbf{A}_k$

$\mathbf{A}_2 = \mathbf{M}_R^{[2]} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \qquad \mathbf{B}_2 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

$\mathbf{A}_3 = \mathbf{M}_R^{[3]} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

$\mathbf{M}_R^{*} = \mathbf{B}_3 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

Transitive Closure

- Towards a more efficient algorithm (Warshall's)

- Definition: Let R be a relation on $S=\{v_1,v_2,\ldots, v_n\}$.
  The *interior vertices* of a path of length m from a to b: $a, x_1, x_2, x_3, \ldots, x_{m-1}, b$ are:

  $x_1, x_2, x_3, \ldots, x_{m-1}$.

Transitive Closure : Method 3

- Warshall's Alg. iteratively constructs 0-1 matrices:

  $\mathbf{W}_0 = \mathbf{M}_R$;

  $\mathbf{W}_1=[w^{[1]}_{ij}]$, where $w^{[1]}_{ij}=1 \leftrightarrow \exists$ a path from v_i to v_j with interior vertices in $\{v_1\}$;

  $\mathbf{W}_2=[w^{[2]}_{ij}]$, where $w^{[2]}_{ij}=1 \leftrightarrow \exists$ a path from v_i to v_j with interior vertices in $\{v_1,v_2\}$…

  $\mathbf{W}_k=[w^{[k]}_{ij}]$, where $w^{[k]}_{ij}=1 \leftrightarrow \exists$ a path from v_i to v_j with interior vertices in $\{v_1,v_2, \ldots, v_k\}$…

  $\mathbf{M}_{R*} = \mathbf{W}_n$.

Transitive Closure – optional

- Lemma: Let R be a relation on $S=\{v_1,v_2,\ldots,v_n\}$, and let $\mathbf{W}_k=[w^{[k]}_{ij}]$ be the 0-1 matrix | $w^{[k]}_{ij}=1 \leftrightarrow \exists$ a path from v_i to v_j with interior vertices in $\{v_1, v_2, \ldots, v_k\}$.

  Then

$$\forall i,j,k \leq n \quad w^{[k]}_{ij} = w^{[k-1]}_{ij} \vee (w^{[k-1]}_{ik} \wedge w^{[k-1]}_{kj}).$$

Warshall's Algorithm: Key Ideas

- In any set A with |A|=n, any transitive relation will be built from a sequence of relations that has a length of at most n. Why?
- Consider the case where the relation *R* on A has the ordered pairs $(a_1,a_2),(a_2,a_3),\ldots,(a_{n-1},a_n)$. Then, (a_1,a_n) must be in *R* for *R* to be transitive
- Thus, by the previous theorem, it suffices to compute (at most) R^n
- Recall that $R^k=R \circ R^{k-1}$ is computed using a bit-matrix product
- The above gives us a natural algorithm for computing the transitive closure: the Warshall's Algorithm

Algorithm 3: Warshall's Algorithm

Input: An (n×n) 0-1 matrix M_R representing a relation R on A, |A|=n

Output: An (n×n) 0-1 matrix W representing the transitive closure of R on A

1. $W \leftarrow M_R$
2. FOR k=1,…, n DO
3. FOR i=1,…,n DO
4. FOR j=1,…,n DO
5. $w_{i,j} \leftarrow w_{i.j} \vee (w_{i,k} \wedge w_{k,j})$
6. END
7. END
8. END
9. RETURN W

// That is, at step *k*: add row *k* to all other rows which have 1 as intersection with *k*th column.

Warshall's Algorithm: Example

- Compute the transitive closure of
  - The relation R={(1,1),(1,2),(1,4),(2,2),(2,3),(3,1),(3,4),(4,1),(4,4)}
  - On the set A={1,2,3,4}

Warshall's Algorithm - Example

- Find the matrix $\mathbf{M}_R$* of the transitive closure of R:

$$\mathbf{M}_R = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

- Solution: $\mathbf{W}_0 = \mathbf{M}_R$

Add row 1 to row 3:

$$\mathbf{W}_1 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & (1) \end{bmatrix}$$

Add row 2 to row 3:

$$\mathbf{W}_2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Add row 3 to row 1:

$$\mathbf{M}_R^* = \mathbf{W}_3 = \begin{bmatrix} 1 & (1) & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Conclusion: transitive closure

- Computing the transitive closure of a digraph is an important problem in many computer science applications:
  - Evaluation of recursive database queries.
  - Analysis of reachability (connectivity) of transition graphs in communication networks.
  - Construction of parsing automata in compilers.