Transitive Closure

Madhavan Mukund

https://www.cmi.ac.in/~madhavan

Mathematics for Data Science 1
Week 11

Transitive closure of a relation

- Let $R \subseteq S \times S$ be a relation on a set S

Transitive closure of a relation

- Let $R \subseteq S \times S$ be a relation on a set S
- For instance, S is a set of people, and $(p, q) \in R$ if p is a parent of q

Transitive closure of a relation

- Let $R \subseteq S \times S$ be a relation on a set S
- For instance, S is a set of people, and $(p, q) \in R$ if p is a parent of q
- Can compute the ancestor relation from the parent relation

Transitive closure of a relation

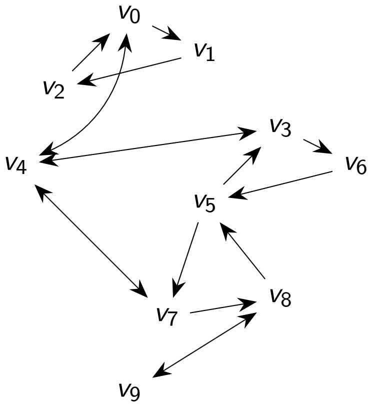
- Let $R \subseteq S \times S$ be a relation on a set S
- For instance, S is a set of people, and $(p, q) \in R$ if p is a parent of q
- Can compute the ancestor relation from the parent relation
- p is an ancestor of q if we can find a sequence of people $r_0, r_1, \ldots, r_n$ such that
  - $p = r_0$
  - For each $i \in \{0, 1, \ldots, n-1\}$, $(r_i, r_{i+1}) \in R$
  - $q = r_n$

Transitive closure of a relation

- Let $R \subseteq S \times S$ be a relation on a set S
- For instance, S is a set of people, and $(p, q) \in R$ if p is a parent of q
- Can compute the ancestor relation from the parent relation
- p is an ancestor of q if we can find a sequence of people $r_0, r_1, \ldots, r_n$ such that
  - $p = r_0$
  - For each $i \in \{0, 1, \ldots, n-1\}$, $(r_i, r_{i+1}) \in R$
  - $q = r_n$
- This is called the transitive closure of R, written R^+
  - $R^+ \subseteq S \times S$ is also a relation
  - R^+ is derived from $R \subseteq S \times S$

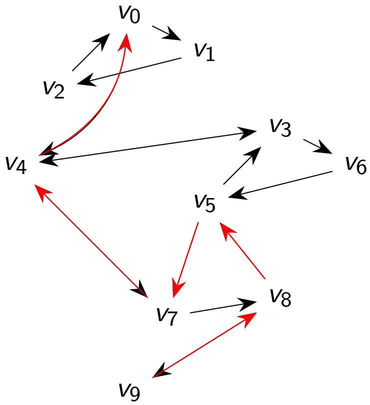
Computing transitive closure

- Represent $R \subseteq S \times S$ as a (directed) graph $G = (V, E)$
  - $V = S$
  - $(u, v) \in E$ if and only if $(u, v) \in R$



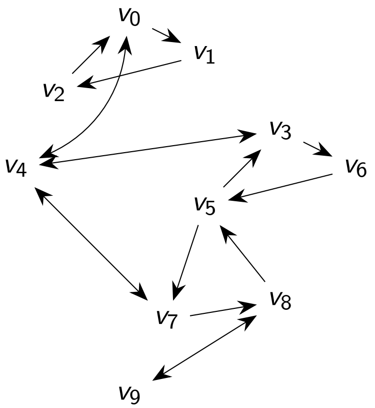
Computing transitive closure

- Represent $R \subseteq S \times S$ as a (directed) graph $G = (V, E)$
  - $V = S$
  - $(u, v) \in E$ if and only if $(u, v) \in R$
- $(u, v) \in R^+$ if and only if there is a path from u to v in the graph



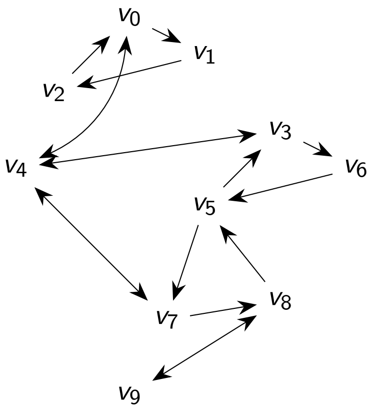
Computing transitive closure

- Represent $R \subseteq S \times S$ as a (directed) graph $G = (V, E)$
  - $V = S$
  - $(u, v) \in E$ if and only if $(u, v) \in R$
- $(u, v) \in R^+$ if and only if there is a path from u to v in the graph
- We know how to compute reachability in graphs
  - BFS, DFS



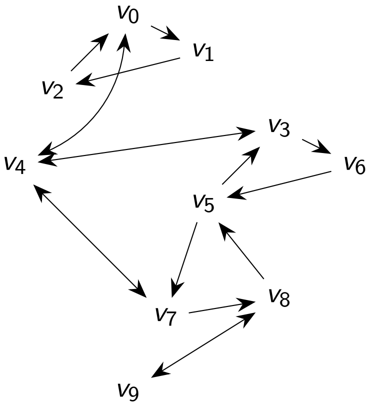
Computing transitive closure

- Represent $R \subseteq S \times S$ as a (directed) graph $G = (V, E)$
  - $V = S$
  - $(u, v) \in E$ if and only if $(u, v) \in R$
- $(u, v) \in R^+$ if and only if there is a path from u to v in the graph
- We know how to compute reachability in graphs
  - BFS, DFS
- Perform BFS/DFS from all verties to compute R^+



Using the adjacency matrix

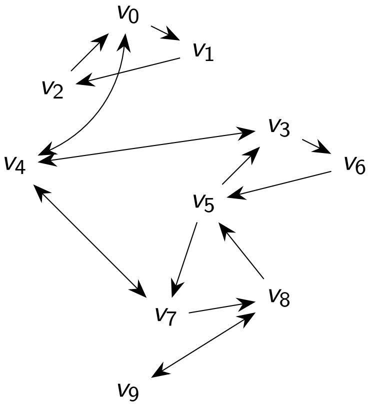
- Another strategy



Using the adjacency matrix

- Another strategy
- Consider the adjacency matrix A for G

	0	1	2	3	4	5	6	7	8	9
0	0	1	0	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	1	0	0	0
4	1	0	0	1	0	0	0	1	0	0
5	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	1	0	0	0	1	0
8	0	0	0	0	0	1	0	0	0	1
9	0	0	0	0	0	0	0	0	1	0

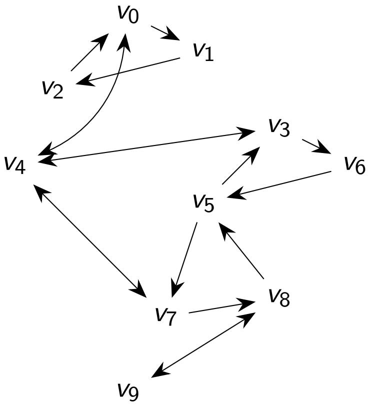


Using the adjacency matrix

- Another strategy
- Consider the adjacency matrix A for G

	0	1	2	3	4	5	6	7	8	9
0	0	1	0	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	1	0	0	0
4	1	0	0	1	0	0	0	1	0	0
5	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	1	0	0	0	1	0
8	0	0	0	0	0	1	0	0	0	1
9	0	0	0	0	0	0	0	0	1	0

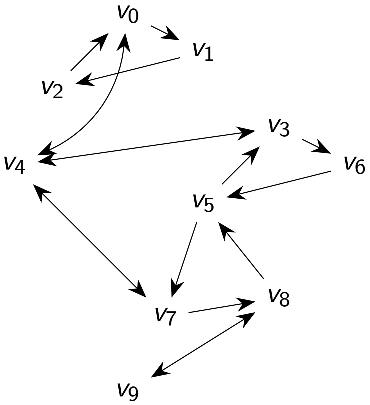
- $A[i,j] = 1$ — path of length 1 from i to j



- Another strategy

- Consider the adjacency matrix A for G

	0	1	2	3	4	5	6	7	8	9
0	0	1	0	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	1	0	0	0
4	1	0	0	1	0	0	0	1	0	0
5	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	1	0	0	0	1	0
8	0	0	0	0	0	1	0	0	0	1
9	0	0	0	0	0	0	0	0	1	0



- $A[i,j] = 1$ — path of length 1 from i to j

- Want $A^+[i,j] = 1$ — path of length ≥ 1 from i to j

Paths of length 2

- Path of length 2 from i to j passes through intermediate k

Paths of length 2

- Path of length 2 from i to j passes through intermediate k
- $A^2[i,j] = 1$ if there is some k such that $A[i,k] = 1$ and $A[k,j] = 1$

A

	0	1	2	3	4	5	6	7	8	9
0	0	1	0	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	1	0	0	0
4	1	0	0	1	0	0	0	1	0	0
5	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	1	0	0	0	1	0
8	0	0	0	0	0	1	0	0	0	1
9	0	0	0	0	0	0	0	0	1	0

A^2

	0	1	2	3	4	5	6	7	8	9
0	1	0	1	1	0	0	0	1	0	0
1										
2										
3										
4										
5										
6										
7										
8										
9										

Paths of length 2

- Path of length 2 from i to j passes through intermediate k
- $A^2[i,j] = 1$ if there is some k such that $A[i,k] = 1$ and $A[k,j] = 1$

A

	0	1	2	3	4	5	6	7	8	9
0	0	1	0	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	1	0	0	0
4	1	0	0	1	0	0	0	1	0	0
5	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	1	0	0	0	1	0
8	0	0	0	0	0	1	0	0	0	1
9	0	0	0	0	0	0	0	0	1	0

A^2

	0	1	2	3	4	5	6	7	8	9
0	1	0	1	1	0	0	0	1	0	0
1	1	0	0	0	0	0	0	0	0	0
2										
3										
4										
5										
6										
7										
8										
9										

Paths of length 2

- Path of length 2 from i to j passes through intermediate k
- $A^2[i,j] = 1$ if there is some k such that $A[i,k] = 1$ and $A[k,j] = 1$

A

	0	1	2	3	4	5	6	7	8	9
0	0	1	0	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	1	0	0	0
4	1	0	0	1	0	0	0	1	0	0
5	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	1	0	0	0	1	0
8	0	0	0	0	0	1	0	0	0	1
9	0	0	0	0	0	0	0	0	1	0

A^2

	0	1	2	3	4	5	6	7	8	9
0	1	0	1	1	0	0	0	1	0	0
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	1	0	0	0	0	0	0
3										
4										
5										
6										
7										
8										
9										

Paths of length 2

- Path of length 2 from i to j passes through intermediate k
- $A^2[i,j] = 1$ if there is some k such that $A[i,k] = 1$ and $A[k,j] = 1$

A

	0	1	2	3	4	5	6	7	8	9
0	0	1	0	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	1	0	0	0
4	1	0	0	1	0	0	0	1	0	0
5	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	1	0	0	0	1	0
8	0	0	0	0	0	1	0	0	0	1
9	0	0	0	0	0	0	0	0	1	0

A^2

	0	1	2	3	4	5	6	7	8	9
0	1	0	1	1	0	0	0	1	0	0
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	1	0	0	0	0	0	0
3										
4										
5										
6										
7										
8										
9										

- Path of length 2 from i to j passes through intermediate k
- $A^2[i,j] = 1$ if there is some k such that $A[i,k] = 1$ and $A[k,j] = 1$

A

	0	1	2	3	4	5	6	7	8	9
0	0	1	0	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	1	0	0	0
4	1	0	0	1	0	0	0	1	0	0
5	0	0	0	1	0	0	0	1	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	1	0	0	0	1	0
8	0	0	0	0	0	1	0	0	0	1
9	0	0	0	0	0	0	0	0	1	0

A^2

	0	1	2	3	4	5	6	7	8	9
0	1	0	1	1	0	0	0	1	0	0
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	1	0	0	0	0	0	0
3	1	0	0	0	0	1	0	1	0	0
4	0	1	0	0	0	0	1	0	1	0
5	0	0	0	0	1	0	1	0	1	0
6	0	0	0	1	0	0	0	1	0	0
7	0	0	0	0	0	1	0	0	0	1
8	0	0	1	0	0	0	0	1	0	0
9	0	0	0	0	0	1	0	0	0	0

Paths of length 3 and beyond

- Extend path of length 2 from i to k by path of length 1 from k to j
- $A^3[i,j] = 1$ if there is some k such that $A^2[i,k] = 1$ and $A[k,j] = 1$

Paths of length 3 and beyond

- Extend path of length 2 from i to k by path of length 1 from k to j
- $A^3[i,j] = 1$ if there is some k such that $A^2[i,k] = 1$ and $A[k,j] = 1$
- Extend path of length 3 from i to k by path of length 1 from k to j
- $A^4[i,j] = 1$ if there is some k such that $A^3[i,k] = 1$ and $A[k,j] = 1$

Paths of length 3 and beyond

- Extend path of length 2 from i to k by path of length 1 from k to j
- $A^3[i,j] = 1$ if there is some k such that $A^2[i,k] = 1$ and $A[k,j] = 1$
- Extend path of length 3 from i to k by path of length 1 from k to j
- $A^4[i,j] = 1$ if there is some k such that $A^3[i,k] = 1$ and $A[k,j] = 1$
- ...
- Extend path of length ℓ from i to k by path of length 1 from k to j
- $A^{\ell+1}[i,j] = 1$ if there is some k such that $A^{\ell}[i,k] = 1$ and $A[k,j] = 1$

Paths of length 3 and beyond

- Extend path of length 2 from i to k by path of length 1 from k to j
- $A^3[i,j] = 1$ if there is some k such that $A^2[i,k] = 1$ and $A[k,j] = 1$
- Extend path of length 3 from i to k by path of length 1 from k to j
- $A^4[i,j] = 1$ if there is some k such that $A^3[i,k] = 1$ and $A[k,j] = 1$
- ...
- Extend path of length ℓ from i to k by path of length 1 from k to j
- $A^{\ell+1}[i,j] = 1$ if there is some k such that $A^{\ell}[i,k] = 1$ and $A[k,j] = 1$
- How long do we go on?

Paths of length 3 and beyond

- Extend path of length 2 from i to k by path of length 1 from k to j
- $A^3[i,j] = 1$ if there is some k such that $A^2[i,k] = 1$ and $A[k,j] = 1$
- Extend path of length 3 from i to k by path of length 1 from k to j
- $A^4[i,j] = 1$ if there is some k such that $A^3[i,k] = 1$ and $A[k,j] = 1$
- ...
- Extend path of length ℓ from i to k by path of length 1 from k to j
- $A^{\ell+1}[i,j] = 1$ if there is some k such that $A^{\ell}[i,k] = 1$ and $A[k,j] = 1$
- How long do we go on?
- Sufficient to check paths upto length $n-1$

Back to transitive closure

- $(i,j) \in R^+$ if there is a path from i to j in G

Back to transitive closure

- $(i, j) \in R^+$ if there is a path from i to j in G
- Length of path is bounded by $n - 1$

Back to transitive closure

- $(i,j) \in R^+$ if there is a path from i to j in G
- Length of path is bounded by $n-1$
- Combine information in $A = A^1, A^2, \ldots, A^{n-1}$ about paths of length 1 to $n-1$

Back to transitive closure

- $(i,j) \in R^+$ if there is a path from i to j in G
- Length of path is bounded by $n-1$
- Combine information in $A = A^1, A^2, \ldots, A^{n-1}$ about paths of length 1 to $n-1$
- $A^+[i,j] = max\{A^\ell[i,j] \mid 1 \leq \ell < n\}$

Back to transitive closure

- $(i,j) \in R^+$ if there is a path from i to j in G
- Length of path is bounded by $n-1$
- Combine information in $A = A^1, A^2, \ldots, A^{n-1}$ about paths of length 1 to $n-1$
- $A^+[i,j] = max\{A^\ell[i,j] \mid 1 \leq \ell < n\}$
  - Each $A^\ell[i,j]$ is either 0 or 1

Back to transitive closure

- $(i,j) \in R^+$ if there is a path from i to j in G
- Length of path is bounded by $n-1$
- Combine information in $A = A^1, A^2, \ldots, A^{n-1}$ about paths of length 1 to $n-1$
- $A^+[i,j] = max\{A^\ell[i,j] \mid 1 \leq \ell < n\}$
  - Each $A^\ell[i,j]$ is either 0 or 1
  - $\max\{A^\ell[i,j] \mid 1 \leq \ell < n\}$ is 1 if and only if $A^\ell[i,j] = 1$ for some $1 \leq \ell < n$

Back to transitive closure

- $(i,j) \in R^+$ if there is a path from i to j in G
- Length of path is bounded by $n-1$
- Combine information in $A = A^1, A^2, \ldots, A^{n-1}$ about paths of length 1 to $n-1$
- $A^+[i,j] = max\{A^\ell[i,j] \mid 1 \leq \ell < n\}$
  - Each $A^\ell[i,j]$ is either 0 or 1
  - $\max\{A^\ell[i,j] \mid 1 \leq \ell < n\}$ is 1 if and only if $A^\ell[i,j] = 1$ for some $1 \leq \ell < n$
  - $A^+[i,j]$ is 1 if and only if there is a path from i to j

Back to transitive closure

- $(i,j) \in R^+$ if there is a path from i to j in G
- Length of path is bounded by $n-1$
- Combine information in $A = A^1, A^2, \ldots, A^{n-1}$ about paths of length 1 to $n-1$
- $A^+[i,j] = max\{A^\ell[i,j] \mid 1 \leq \ell < n\}$
  - Each $A^\ell[i,j]$ is either 0 or 1
  - $\max\{A^\ell[i,j] \mid 1 \leq \ell < n\}$ is 1 if and only if $A^\ell[i,j] = 1$ for some $1 \leq \ell < n$
  - $A^+[i,j]$ is 1 if and only if there is a path from i to j
- This calculation can be described directly using matrix multiplication

Back to transitive closure

- $(i,j) \in R^+$ if there is a path from i to j in G
- Length of path is bounded by $n-1$
- Combine information in $A = A^1, A^2, \ldots, A^{n-1}$ about paths of length 1 to $n-1$
- $A^+[i,j] = max\{A^\ell[i,j] \mid 1 \leq \ell < n\}$
  - Each $A^\ell[i,j]$ is either 0 or 1
  - $\max\{A^\ell[i,j] \mid 1 \leq \ell < n\}$ is 1 if and only if $A^\ell[i,j] = 1$ for some $1 \leq \ell < n$
  - $A^+[i,j]$ is 1 if and only if there is a path from i to j
- This calculation can be described directly using matrix multiplication
- $A^2 = A \times A$, $A^3 = A^2 \times A$, $\ldots$, $A^{\ell+1} = A^\ell \times A$

Summary

- The transitive closure $R+$ of a relation R connects pairs of elements related by a sequence of intermediate elements

Summary

- The transitive closure $R+$ of a relation R connects pairs of elements related by a sequence of intermediate elements
- A typical example is the ancestor relation derived from the parent relation

Summary

- The transitive closure $R+$ of a relation R connects pairs of elements related by a sequence of intermediate elements
- A typical example is the ancestor relation derived from the parent relation
- If we represent a relation as a graph, transitive closure corresponds to reachability

Summary

- The transitive closure $R+$ of a relation R connects pairs of elements related by a sequence of intermediate elements
- A typical example is the ancestor relation derived from the parent relation
- If we represent a relation as a graph, transitive closure corresponds to reachability
- Reachability between all pairs of vertices can be checked using repeated BFS/DFS starting from each vertex

Summary

- The transitive closure $R+$ of a relation R connects pairs of elements related by a sequence of intermediate elements
- A typical example is the ancestor relation derived from the parent relation
- If we represent a relation as a graph, transitive closure corresponds to reachability
- Reachability between all pairs of vertices can be checked using repeated BFS/DFS starting from each vertex
- Alternatively, we can perform repeated matrix multiplication on the adjacency matrix A, observing that the length of a path is at most $n-1$