COMP90042

Workshop Week 04

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

Syllabus

1	Introduction and Preprocessing	Text classification
2	Lexical semantics	Distributional semantics
3	Part of Speech Tagging	Hidden Markov Models
4	Probabilistic Sequence Modelling	Context-Free Grammars
5	Probabilistic Parsing	Dependency parsing
	Easter holiday break	
6	N-gram language modelling	Deep learning for language models and tagging
7	Information Extraction	Question Answering
8	Topic Models	*ANZAC day holiday*
9	Information Retrieval -- Boolean search and the vector space model	Indexing and querying in the vector space model, evaluation
10	Index and vocabulary compression	Efficient query processing
11	The Web as a Graph: Page-rank & HITS	Machine Translation (word based)
12	Machine translation (phrase based) and neural encoder-decoder	Subject review

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

Outline

- Part of Speech (POS) tags
  - Examples
- Hidden Markov Models (HMMs)
  - Directed Probabilistic Graphical Models
  - A dynamic programming example
  - The Viterbi algorithm

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

- POS tag (by hand) the following sentence:
- Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.

Pierre	the
Vinken	board
,	as
61	a
years	nonexecutive
old	director
,	Nov.
will	29
join	.

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

Major Penn Treebank tags

NN noun

JJ adjective

DT determiner

IN preposition

MD modal

RP particle

TO *to*

VB verb

RB adverb

CD cardinal number

PRP personal pronoun

CC coordinating conjunction

WH wh-pronoun

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

Penn treebank derived tags

NN: NNS (plural, *wombats*), NNP (proper, *Australia*), NNPS (proper plural, *Australians*)

VB: VBP (base, *eat*), VB (infinitive, *eat*), VBZ (3rd person singular, *eats*), VBD (past tense, *ate*), VBG (gerund, *eating*), VBN (past participle, *eaten*)

JJ: JJR (comparative, *nicer*), JJS (superlative, *nicest*)

RB: RBR (comparative, *faster*), RBS (superlative, *fastest*)

PRP: PRP$ (possessive, *my*)

WH: WH$ (possessive, *whose*), WDT(*wh*-determiner, *who*), WRB (*wh*-adverb, *where*)

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

- POS tag (by hand) the following sentence:
- Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.

Pierre	the
Vinken	board
,	as
61	a
years	nonexecutive
old	director
,	Nov.
will	29
join	.

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

- POS tag (by hand) the following sentence:
- Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.

NNP	Pierre	DT	the
NNP	Vinken	NN	board
,	,	IN	as
CD	61	DT	a
NNS	years	JJ	nonexecutive
JJ	old	NN	director
,	,	NNP	Nov.
MD	will	CD	29
VB	join	.	.

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

Outline

- Part of Speech (POS) tags
  - Examples
- Hidden Markov Models (HMMs)
  - Directed Probabilistic Graphical Models
  - A dynamic programming example
  - The Viterbi algorithm

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

The formula for directed PGMs

- Directed PGMs
  - = Directed Probabilistic Graphical Models

$$P(\text{all } r.v.) = \prod_{\text{every } r.v.} P(r.v.\,|\text{parents of } r.v.)$$

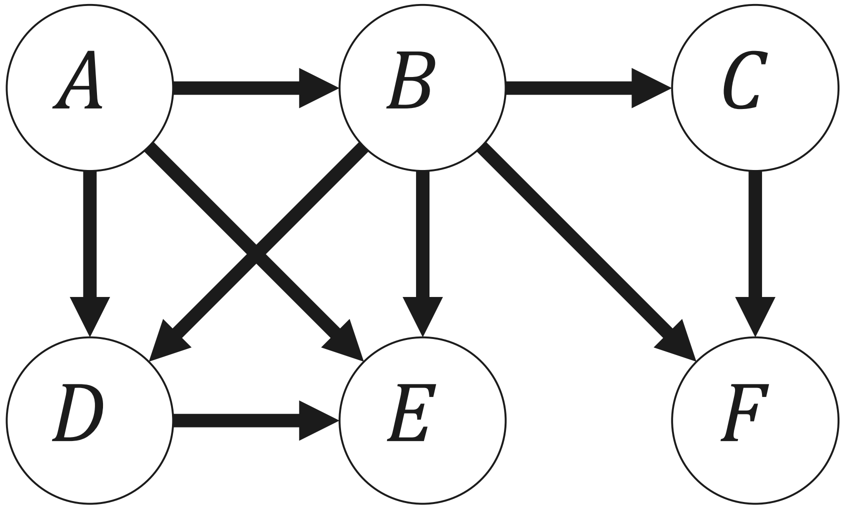
COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

The formula for directed PGMs

- Directed PGMs
  - = Directed Probabilistic Graphical Models

$$P(\text{all } r.v.) = \prod_{\text{every } r.v.} P(r.v. \mid \text{parents of } r.v.)$$

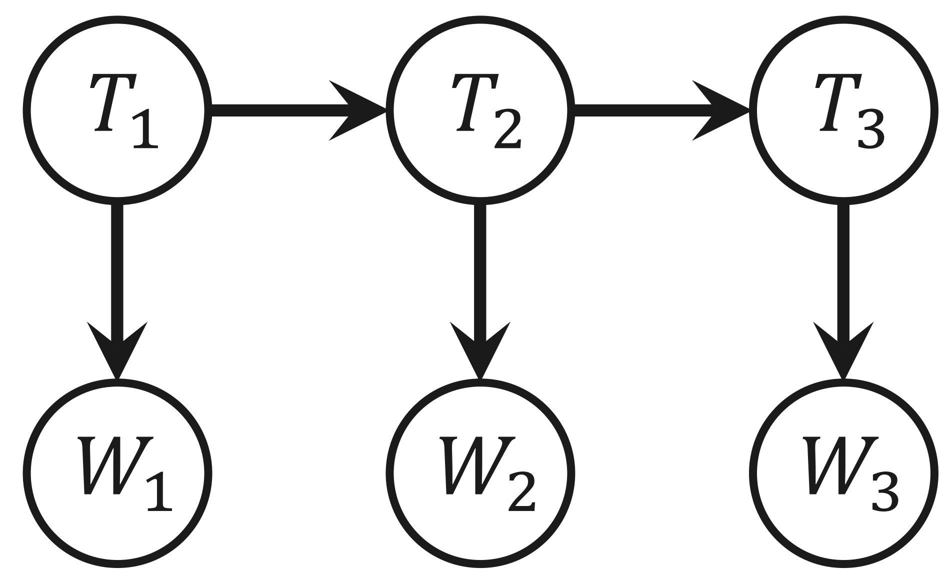
$$\begin{aligned} & P(A,B,C,D,E,F) \\ = & P(A)P(B|A)P(C|B) \\ \cdot & P(D|A,B)P(E|A,B,D)P(F|B,C) \end{aligned}$$



COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

HMMs are directed PGMs

- W_1, W_2, W_3 are words in a sentence
- T_1, T_2, T_3 are their POS tags (random variables)
- We define the joint distribution as follows:

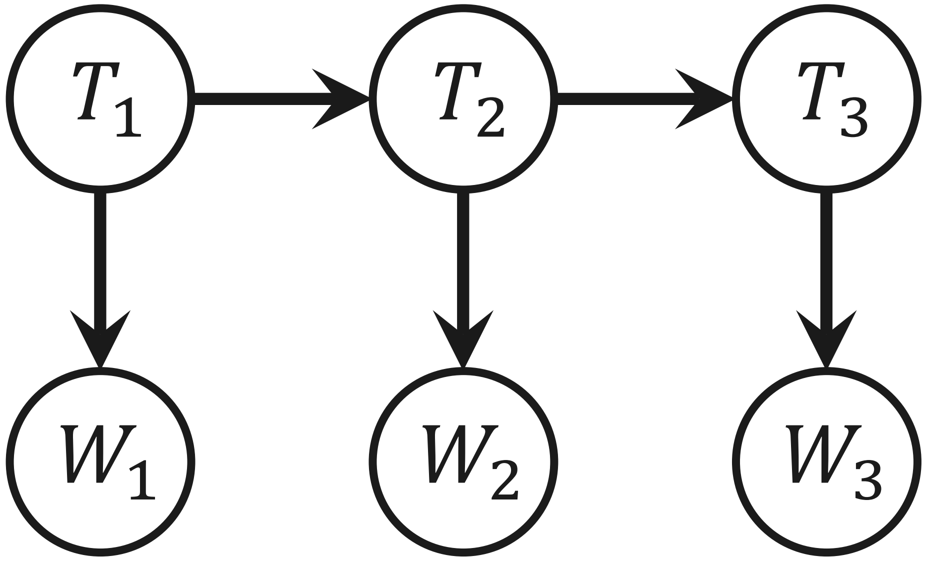


$$P(T_1, T_2, T_3, W_1, W_2, W_3) = \begin{matrix} P(T_1)P(T_2|T_1)P(T_3|T_2) \\ P(W_1|T_1)P(W_2|T_2)P(W_3|T_3) \end{matrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

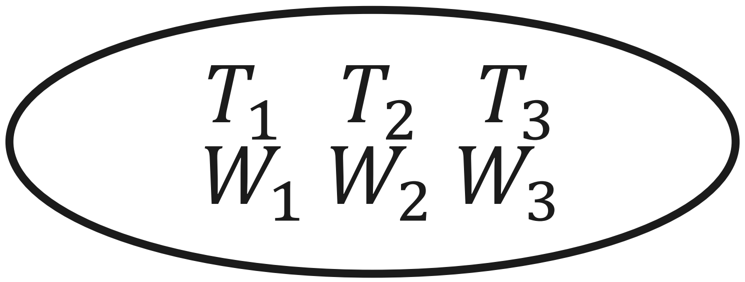
HMMs are directed PGMs

- W_1, W_2, W_3 are words in a sentence
- T_1, T_2, T_3 are their POS tags (random variables)
- We define the joint distribution as follows:
- But why the graph is like this?



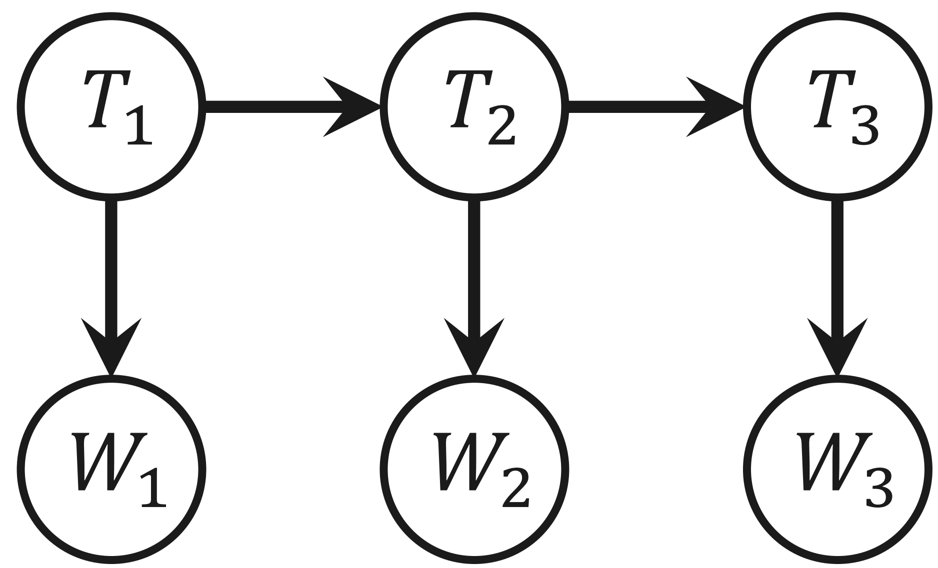
$$P(T_1, T_2, T_3, W_1, W_2, W_3) = \begin{matrix} P(T_1)P(T_2|T_1)P(T_3|T_2) \\ P(W_1|T_1)P(W_2|T_2)P(W_3|T_3) \end{matrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE



The original graph
(without any assumptions)

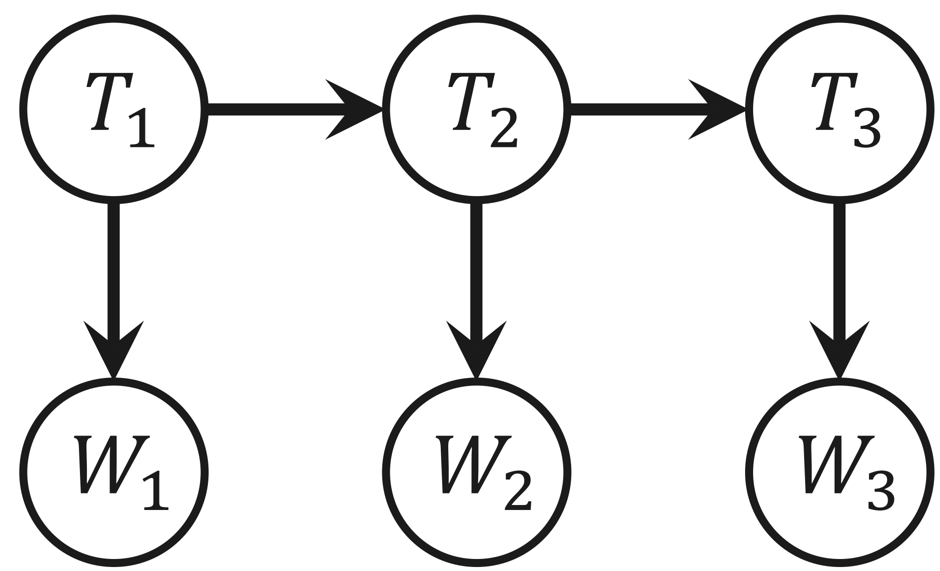
$$P(T_1, T_2, T_3, W_1, W_2, W_3)$$



$$P(T_1)P(T_2|T_1)P(T_3|T_2)$$
$$P(W_1|T_1)P(W_2|T_2)P(W_3|T_3)$$

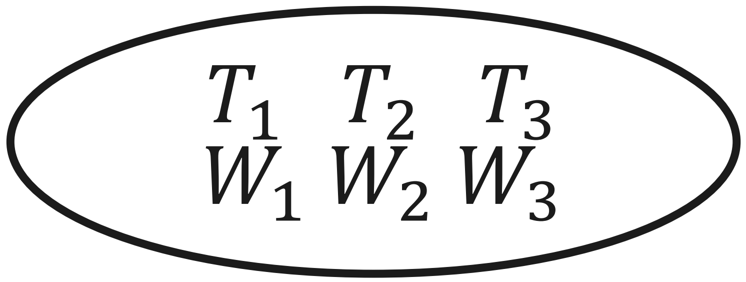
COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$P(T_1, T_2, T_3, W_1, W_2, W_3) = P(T_1, T_2, T_3)P(W_1, W_2, W_3|T_1, T_2, T_3)$$

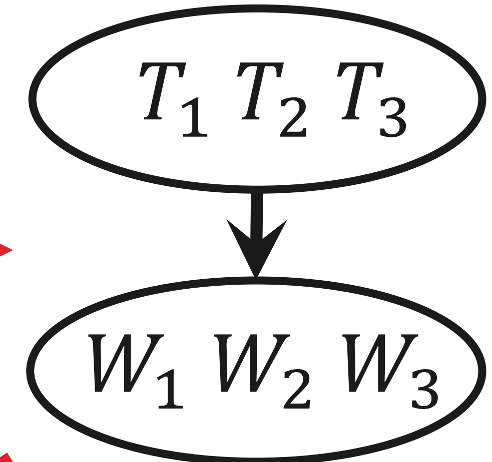


$$P(T_1)P(T_2|T_1)P(T_3|T_2)$$
$$P(W_1|T_1)P(W_2|T_2)P(W_3|T_3)$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

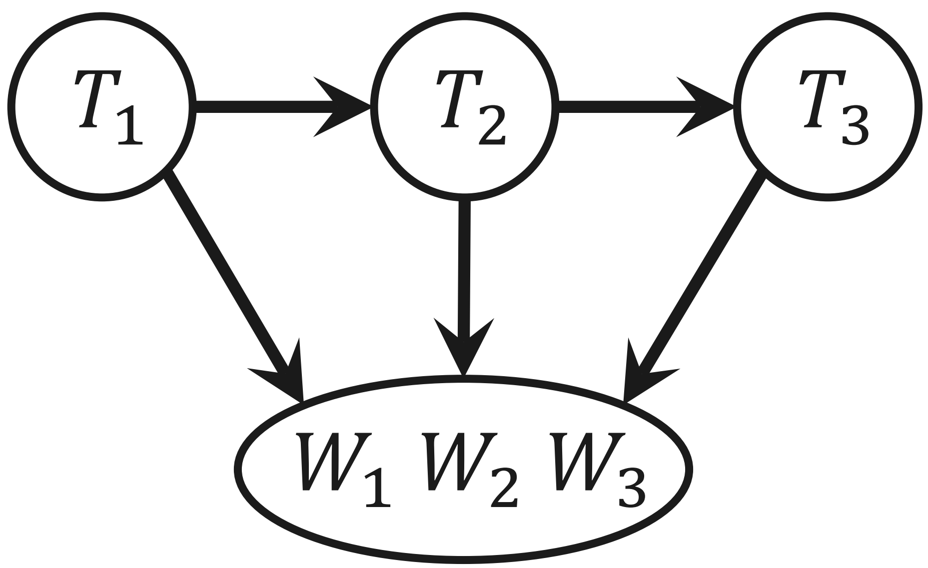


always holds

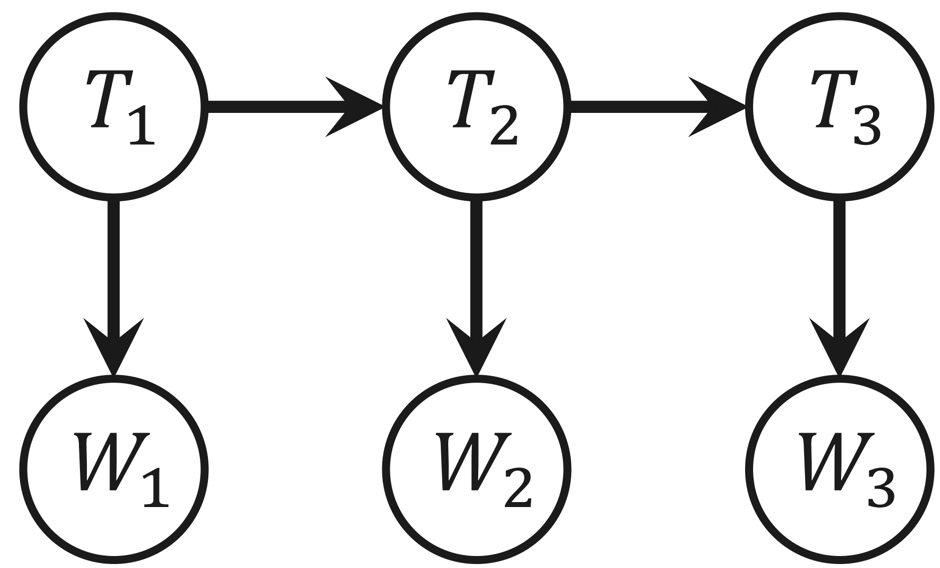


Markov assumption

$$P(T_1, T_2, T_3, W_1, W_2, W_3) = P(T_1, T_2, T_3)P(W_1, W_2, W_3 | T_1, T_2, T_3)$$

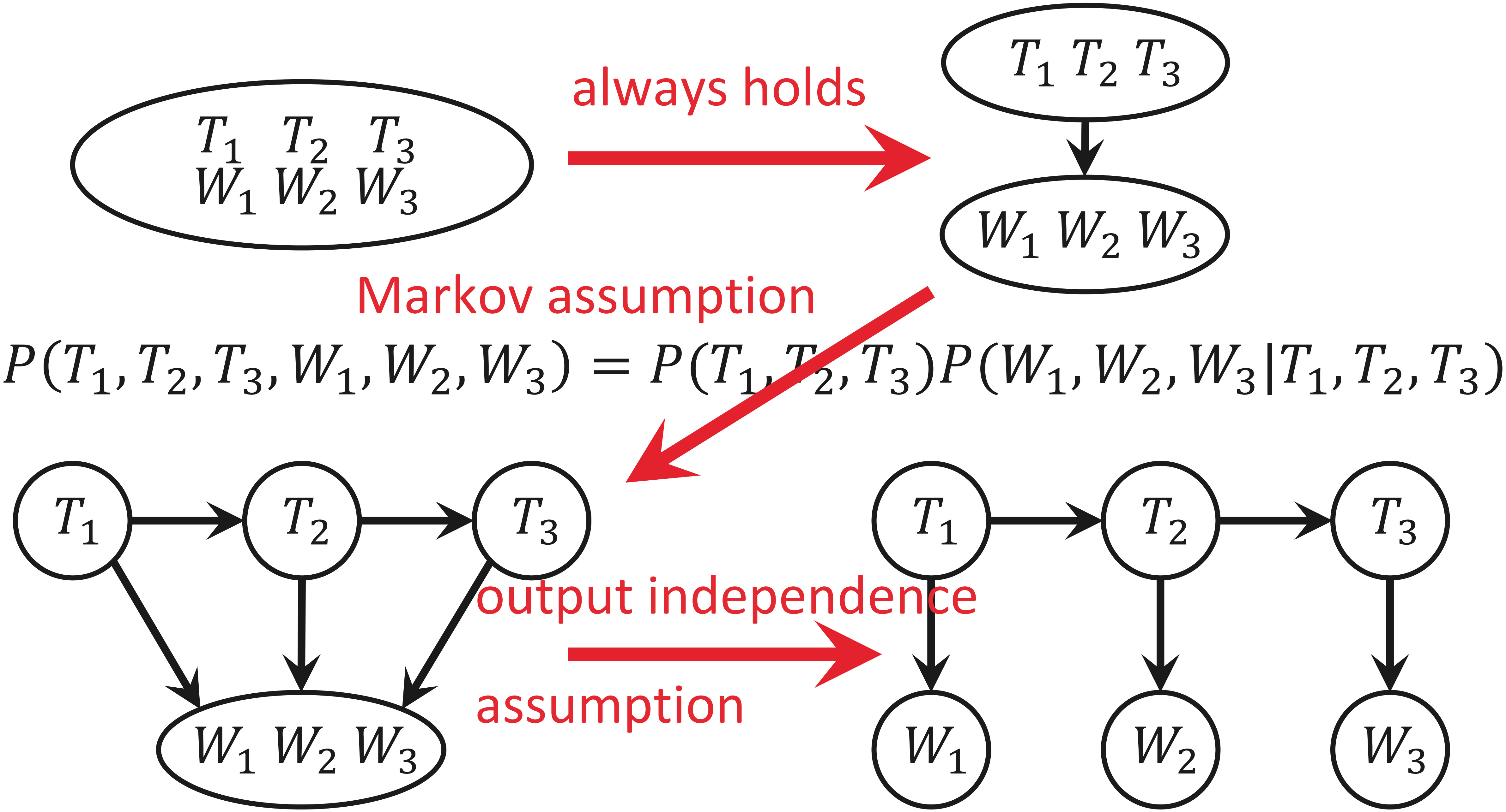


$$P(T_1)P(T_2|T_1)P(T_3|T_2)$$
$$P(W_1, W_2, W_3|T_1, T_2, T_3)$$



$$P(T_1)P(T_2|T_1)P(T_3|T_2)$$
$$P(W_1|T_1)P(W_2|T_2)P(W_3|T_3)$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE



$$P(T_1, T_2, T_3, W_1, W_2, W_3) = P(T_1, T_2, T_3)P(W_1, W_2, W_3|T_1, T_2, T_3)$$

$$P(T_1)P(T_2|T_1)P(T_3|T_2)$$
$$P(W_1, W_2, W_3|T_1, T_2, T_3)$$

$$P(T_1)P(T_2|T_1)P(T_3|T_2)$$
$$P(W_1|T_1)P(W_2|T_2)P(W_3|T_3)$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

A summary

- *Markov assumption*: Only fixed number k of recent tags are relevant (k is known as the *Markov order;* in above $k=1$)
  - $P(T_1, T_2, T_3) = P(T_1)P(T_2|T_1)P(T_3|T_2)$

- *Limited dependency between words and their tags*: Tags are assumed to capture the local context needed to explain the observations
  - $P(W_1, W_2, W_3|T_1, T_2, T_3) = P(W_1|T_1)P(W_2|T_2)P(W_3|T_3)$

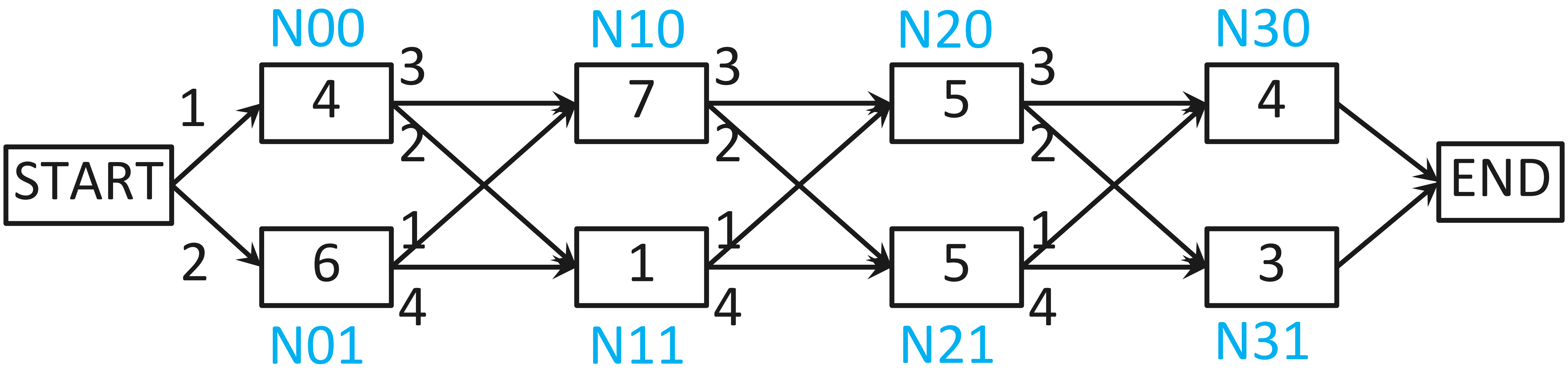
COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

Outline

- Part of Speech (POS) tags
  - Examples
- Hidden Markov Models (HMMs)
  - Directed Probabilistic Graphical Models
  - A dynamic programming example
  - The Viterbi algorithm

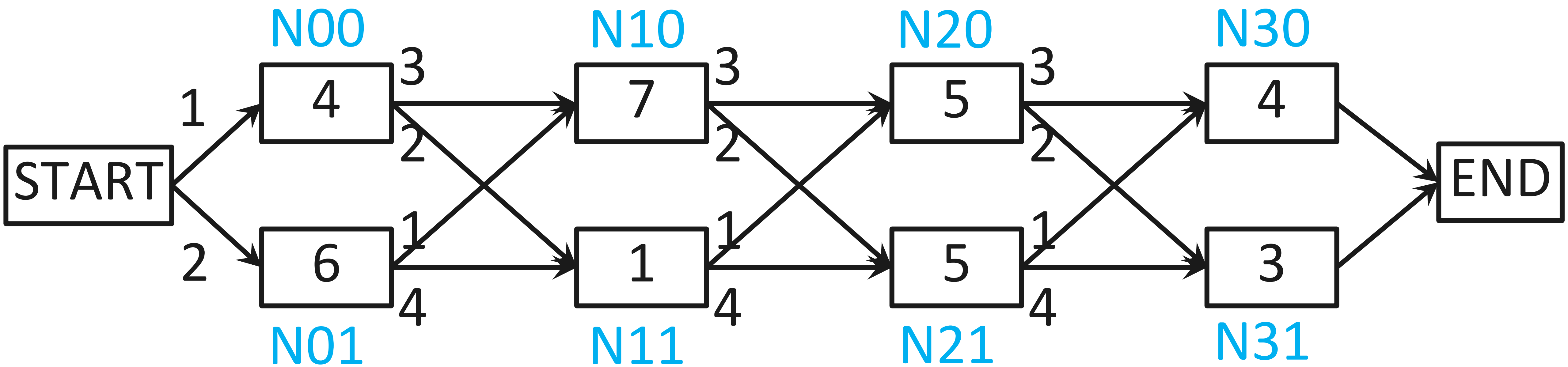
COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

Max reward problem



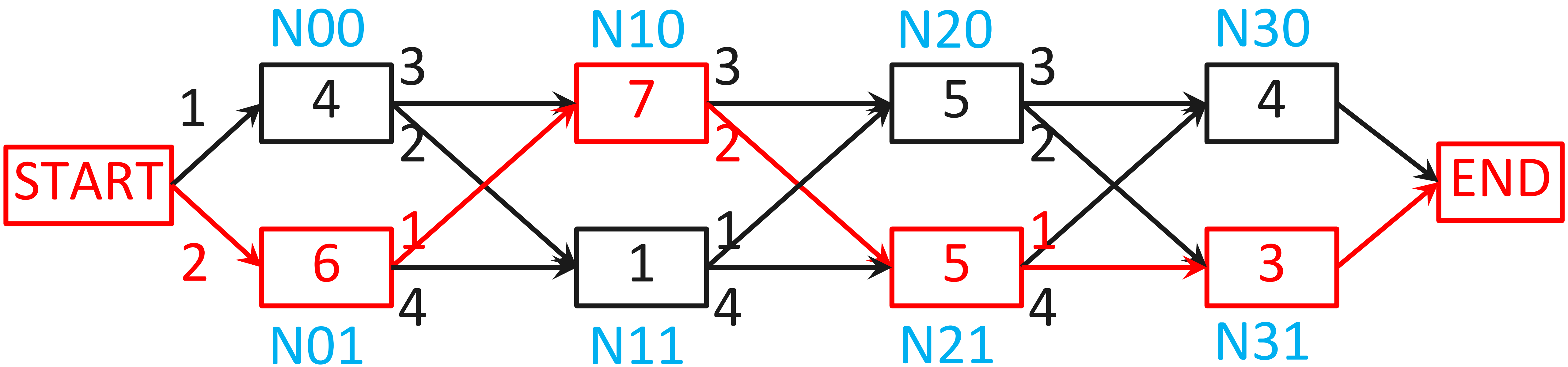
COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix}1\\2\end{bmatrix}, A = \begin{bmatrix}3 & 2\\1 & 4\end{bmatrix}, B = \begin{bmatrix}4 & 7 & 5 & 4\\6 & 1 & 5 & 3\end{bmatrix}$$



COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

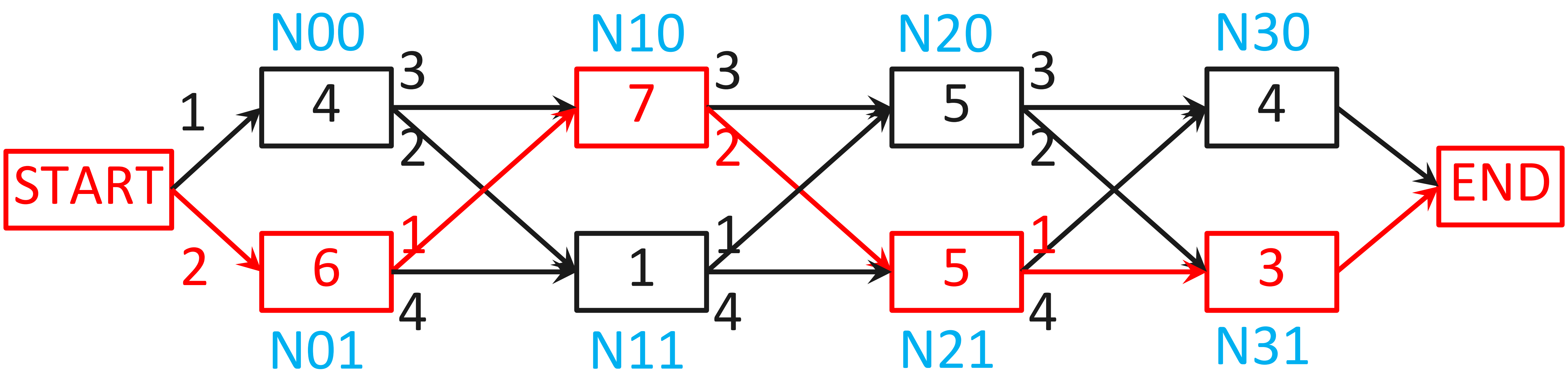
$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$



- Reward $= 2 + 6 + 1 + 7 + 2 + 5 + 4 + 3 = 30$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix}1\\2\end{bmatrix}, A = \begin{bmatrix}3 & 2\\1 & 4\end{bmatrix}, B = \begin{bmatrix}4 & 7 & 5 & 4\\6 & 1 & 5 & 3\end{bmatrix}$$



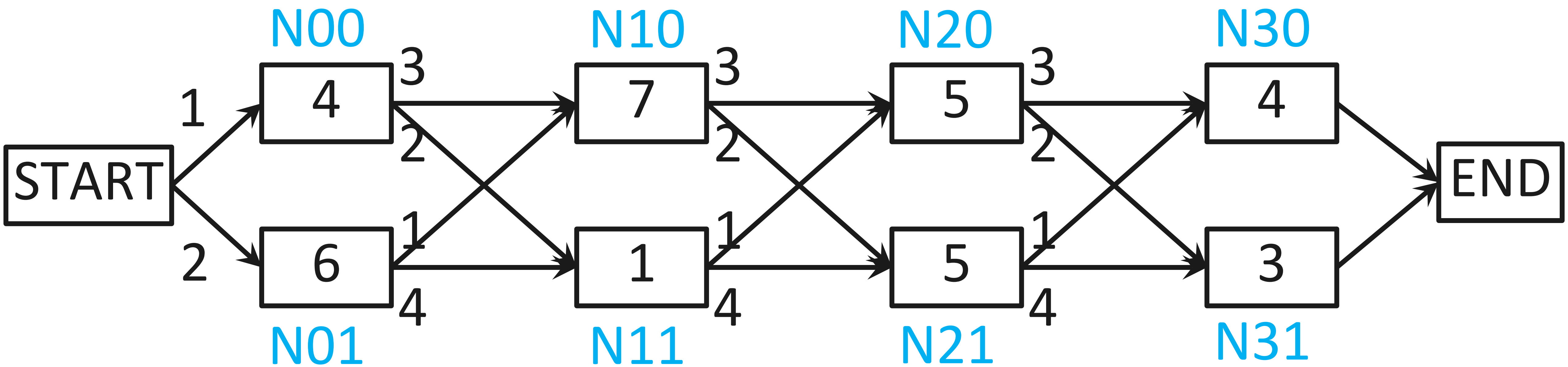
- Reward $= 2 + 6 + 1 + 7 + 2 + 5 + 4 + 3 = 30$

```
seq = [1, 0, 1, 1]

reward = pi[seq[0]] + B[seq[0], 0]
for i in [1, 2, 3]:
    reward += A[seq[i-1], seq[i]] + B[seq[i], i]
```

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$

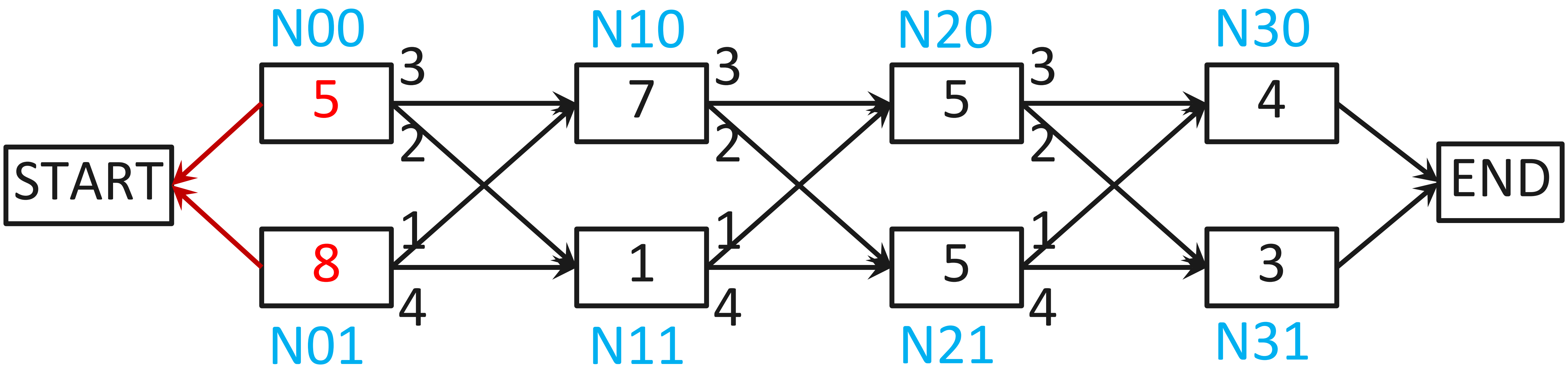


- Best path from S to N00: S->N00 (1+4=5)
- Best path from S to N01: S->N01 (2+6=8)

$$alpha = \begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \end{bmatrix}, backp = \begin{bmatrix} ? & ? & ? & ? \\ ? & ? & ? & ? \end{bmatrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$

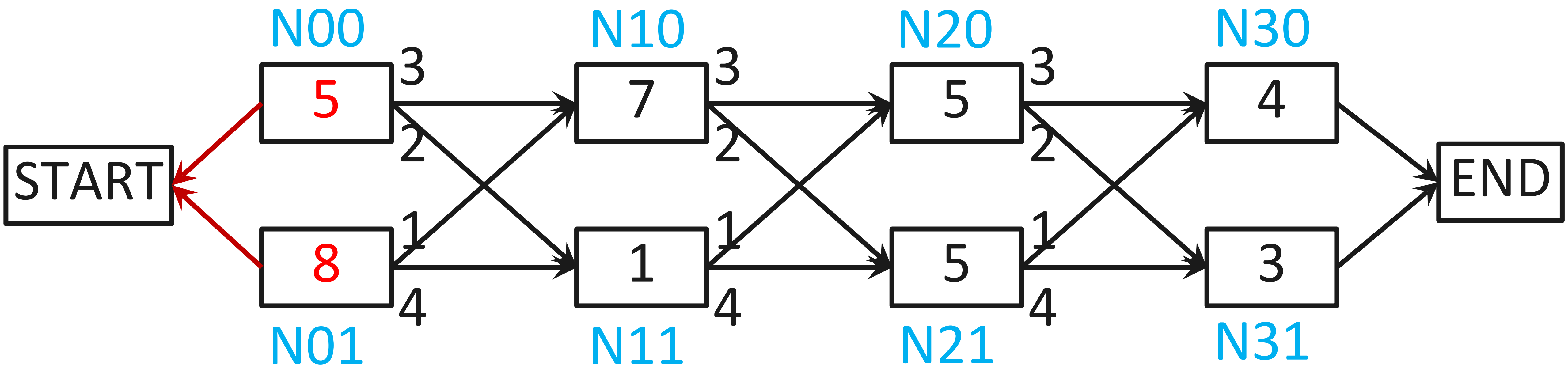


- Best path from S to N00: S->N00 (1+4=5)
- Best path from S to N01: S->N01 (2+6=8)

$$alpha = \begin{bmatrix} 5 & ? & ? & ? \\ 8 & ? & ? & ? \end{bmatrix}, backp = \begin{bmatrix} 0 & ? & ? & ? \\ 0 & ? & ? & ? \end{bmatrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$

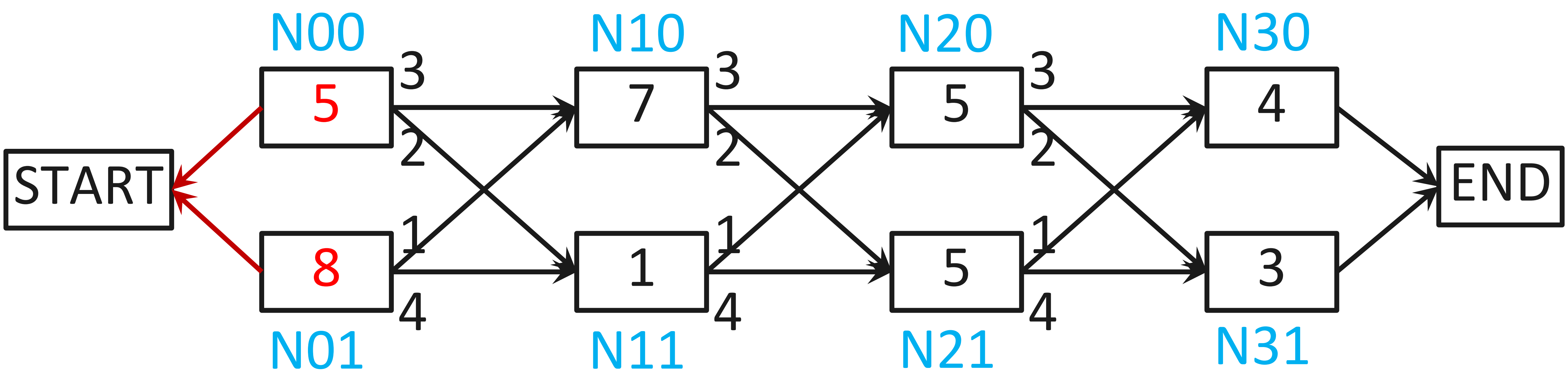


- Best path from S to N10: ?
- Best path from S to N11: ?

$$alpha = \begin{bmatrix} 5 & ? & ? & ? \\ 8 & ? & ? & ? \end{bmatrix}, backp = \begin{bmatrix} 0 & ? & ? & ? \\ 0 & ? & ? & ? \end{bmatrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

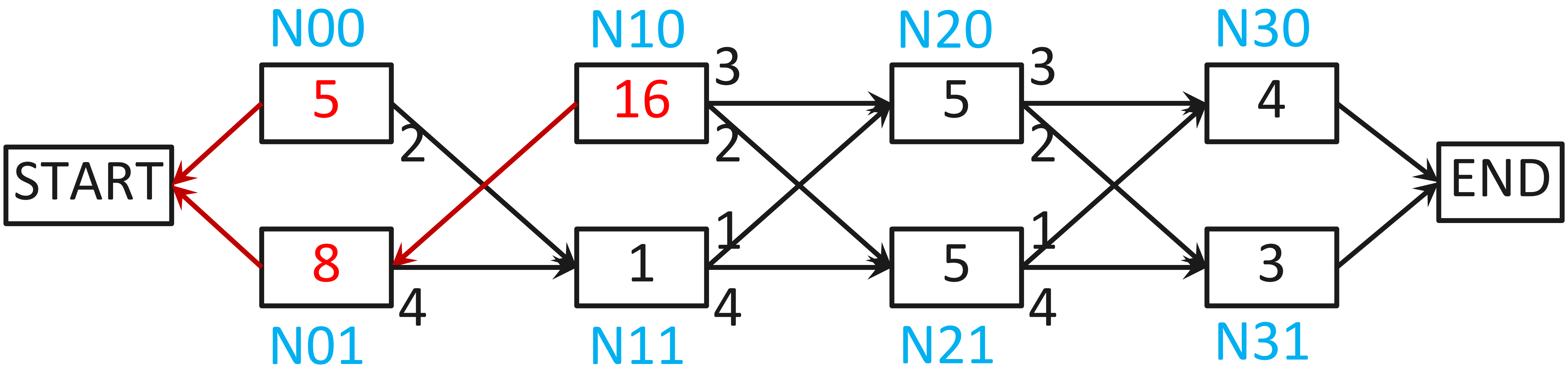
$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$



- Best path from S to N10: ?
  - S->N00->N10 = 5 + 3 + 7 = 15
  - S->N01->N10 = 8 + 1 + 7 = 16
- Best path from S to N11: ?

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

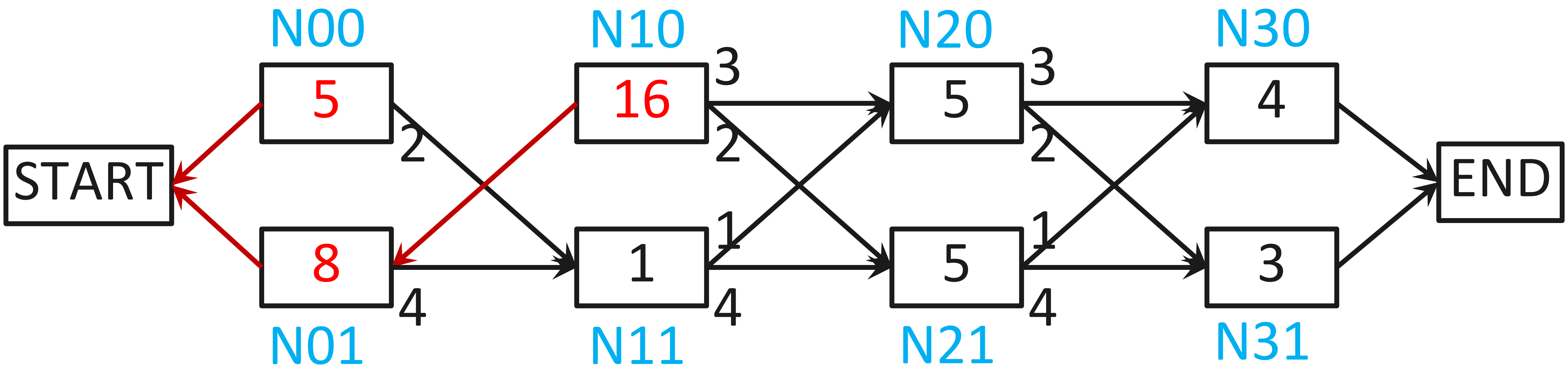
$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$



- Best path from S to N10: ?
  - S->N00->N10 = 5 + 3 + 7 = 15
  - S->N01->N10 = 8 + 1 + 7 = 16
- Best path from S to N11: ?

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

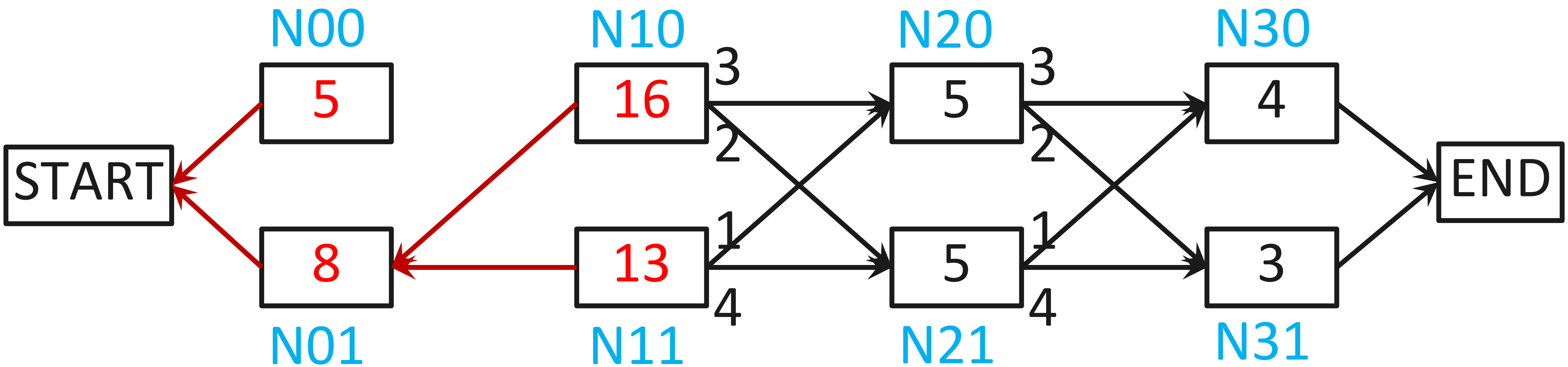
$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$



- Best path from S to N10: S->N01->N10 (8+1+7=16)
- Best path from S to N11: ?
  - S->N00->N11 = 5 + 2 + 1 = 8
  - S->N01->N11 = 8 + 4 + 1 = 13

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$

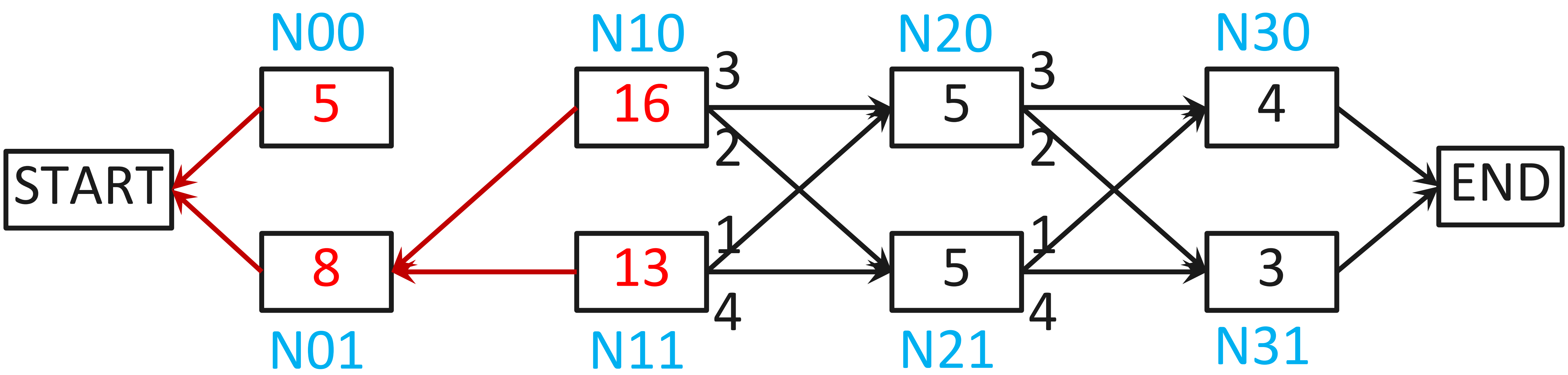


- Best path from S to N10: S->N01->N10 (8+1+7=16)
- Best path from S to N11: S->N01->N11 (8+4+1=13)

$$alpha = \begin{bmatrix} 5 & 16 & ? & ? \\ 8 & 13 & ? & ? \end{bmatrix}, backp = \begin{bmatrix} 0 & 1 & ? & ? \\ 0 & 1 & ? & ? \end{bmatrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$

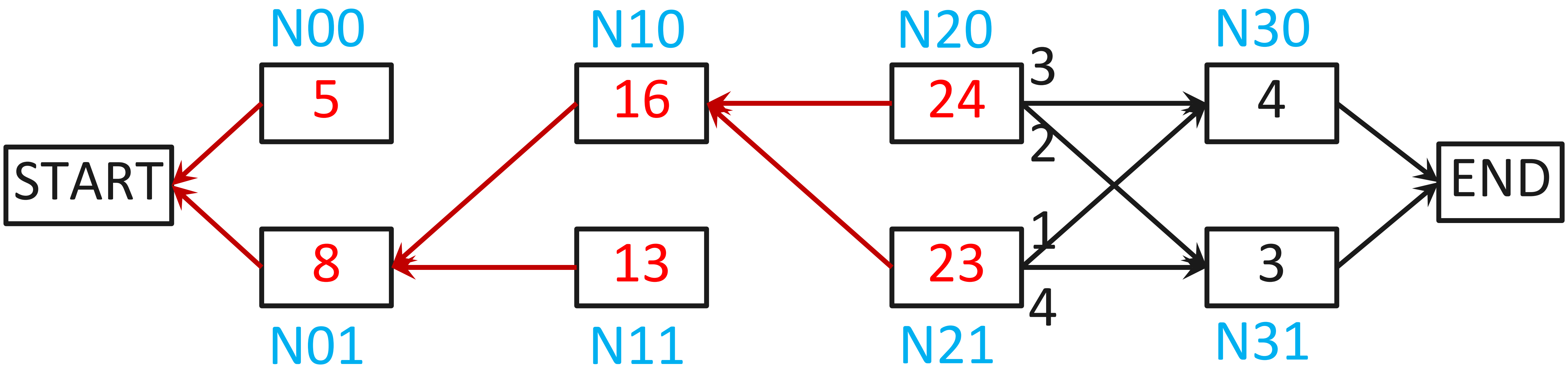


- Best path from S to N20: ?
- Best path from S to N21: ?

$$alpha = \begin{bmatrix} 5 & 16 & ? & ? \\ 8 & 13 & ? & ? \end{bmatrix}, backp = \begin{bmatrix} 0 & 1 & ? & ? \\ 0 & 1 & ? & ? \end{bmatrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$

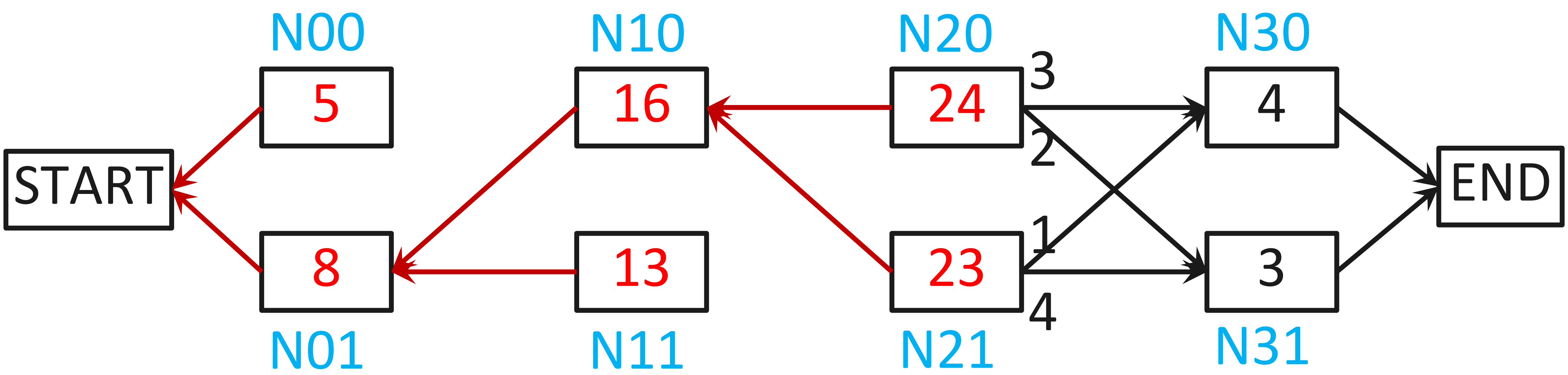


- Best path from S to N20: S->N01->N10->N20 (24)
- Best path from S to N21: S->N01->N10->N21 (23)

$$alpha = \begin{bmatrix} 5 & 16 & 24 & ? \\ 8 & 13 & 23 & ? \end{bmatrix}, backp = \begin{bmatrix} 0 & 1 & 0 & ? \\ 0 & 1 & 0 & ? \end{bmatrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$

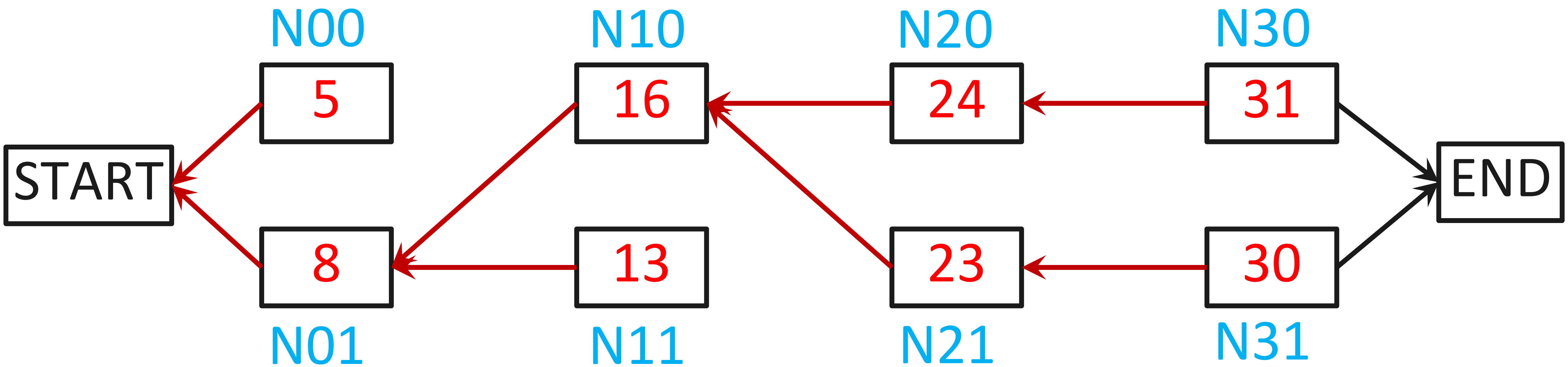


- Best path from S to N30: ?
- Best path from S to N31: ?

$$alpha = \begin{bmatrix} 5 & 16 & 24 & ? \\ 8 & 13 & 23 & ? \end{bmatrix}, backp = \begin{bmatrix} 0 & 1 & 0 & ? \\ 0 & 1 & 0 & ? \end{bmatrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$

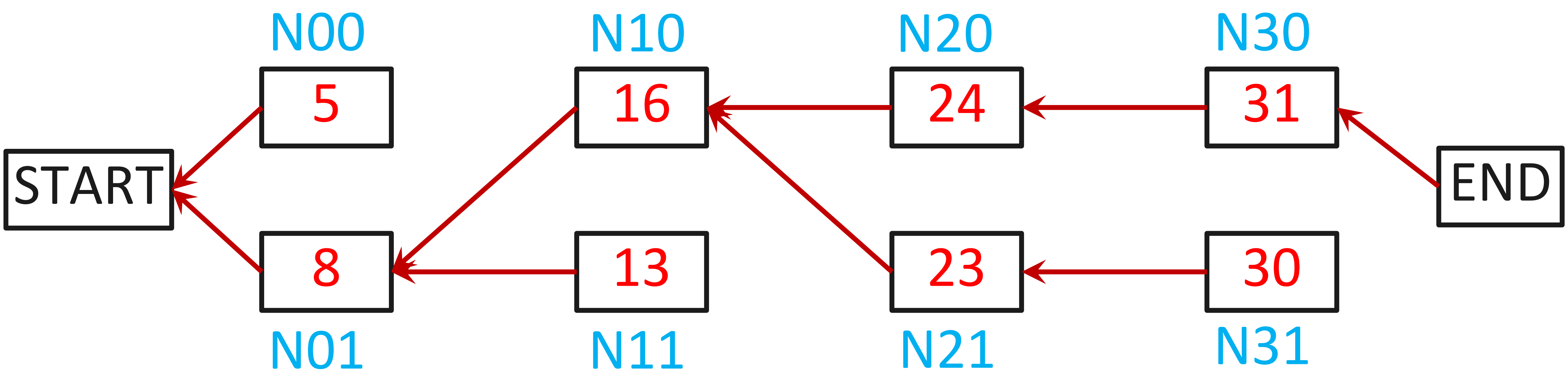


- Best path from S to N30: S->N01->N10->N20->N30 (31)
- Best path from S to N31: S->N01->N10->N21->N31 (30)

$$alpha = \begin{bmatrix} 5 & 16 & 24 & 31 \\ 8 & 13 & 23 & 30 \end{bmatrix}, backp = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

$$\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}, B = \begin{bmatrix} 4 & 7 & 5 & 4 \\ 6 & 1 & 5 & 3 \end{bmatrix}$$



- Best path from S to E: S->N01->N10->N20->N30->E (31)

$$alpha = \begin{bmatrix} 5 & 16 & 24 & 31 \\ 8 & 13 & 23 & 30 \end{bmatrix}, backp = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE

Outline

- Part of Speech (POS) tags
  - Examples
- Hidden Markov Models (HMMs)
  - Directed Probabilistic Graphical Models
  - A dynamic programming example
  - The Viterbi algorithm
    - See the details in the notebook

COPYRIGHT 2018, THE UNIVERSITY OF MELBOURNE