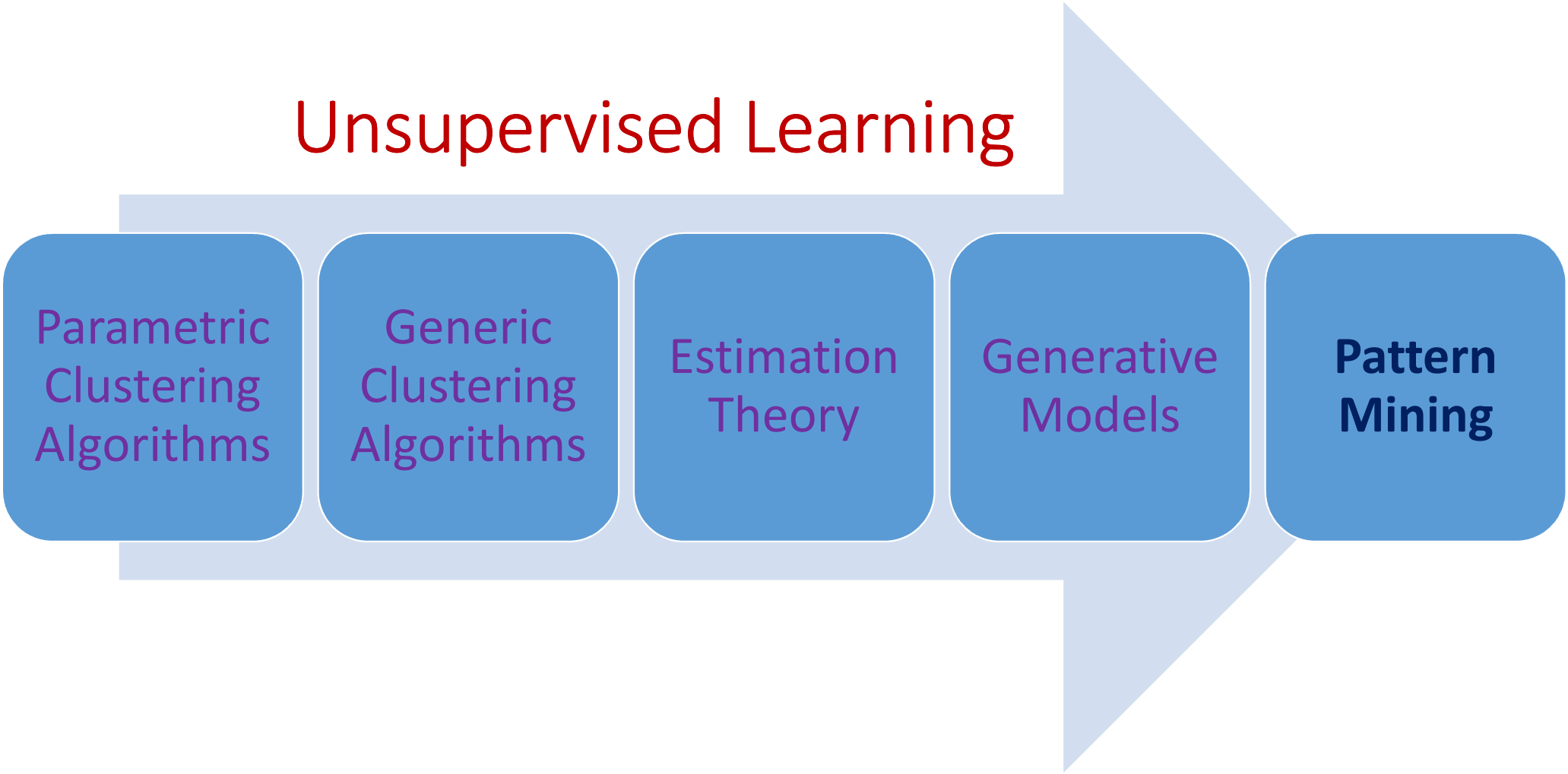
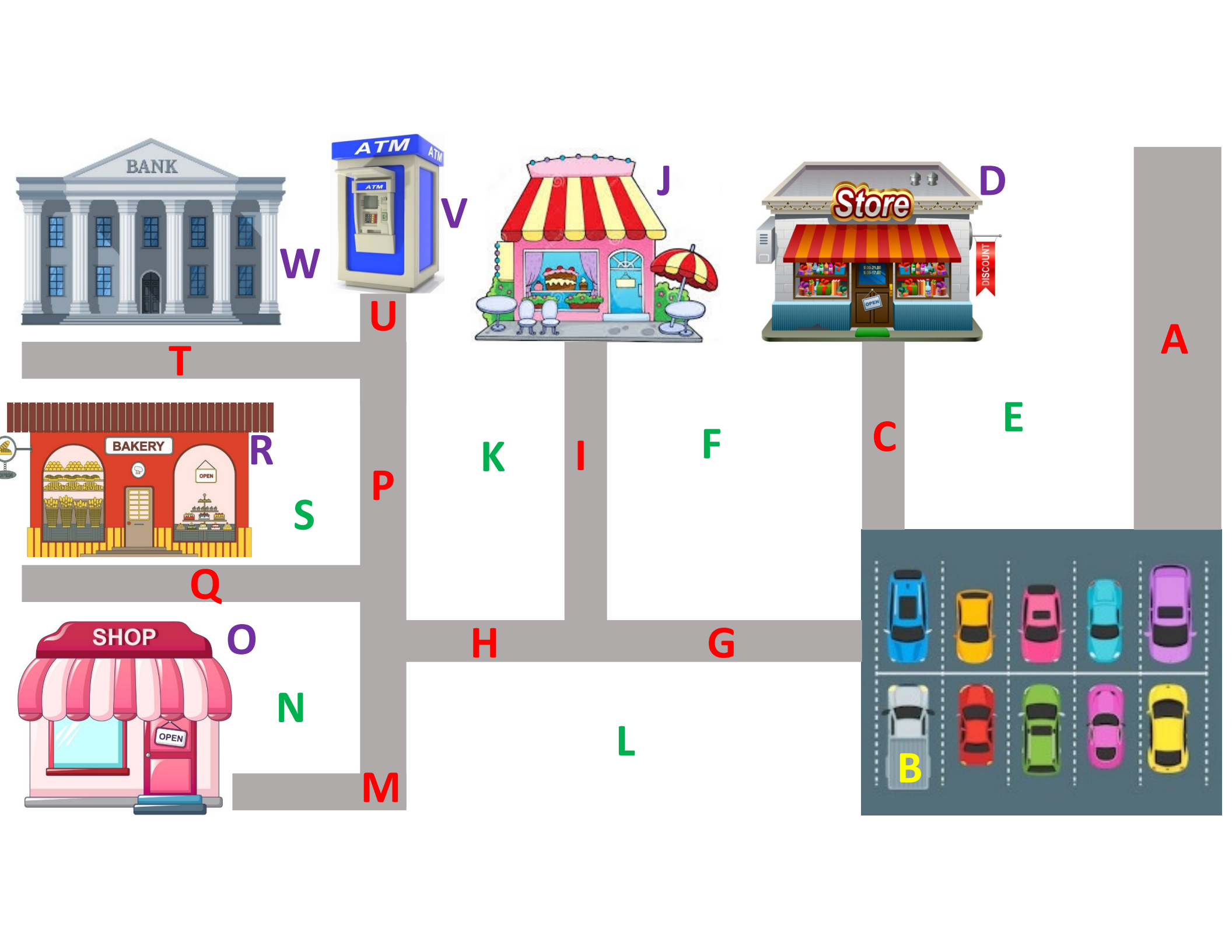
Learning Suffix Trees

Prithwijit Guha
Dept. of EEE, IIT Guwahati

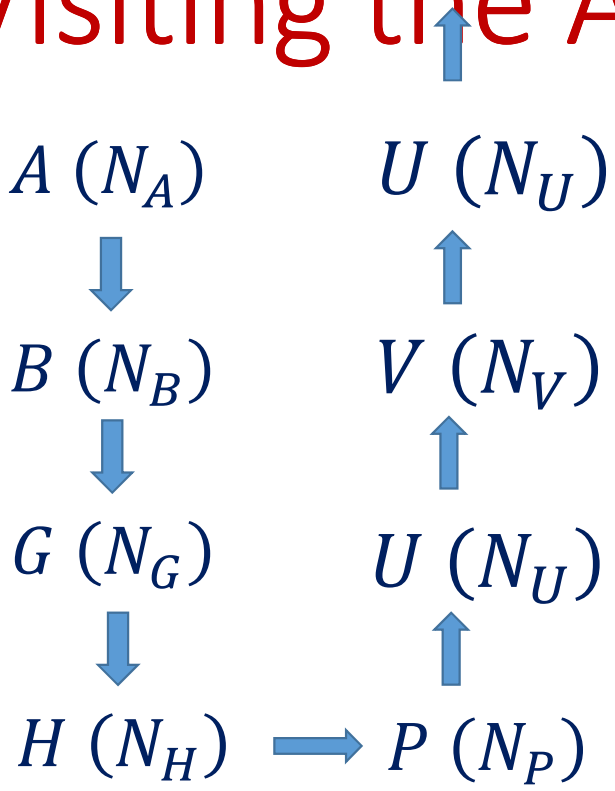
Unsupervised Learning

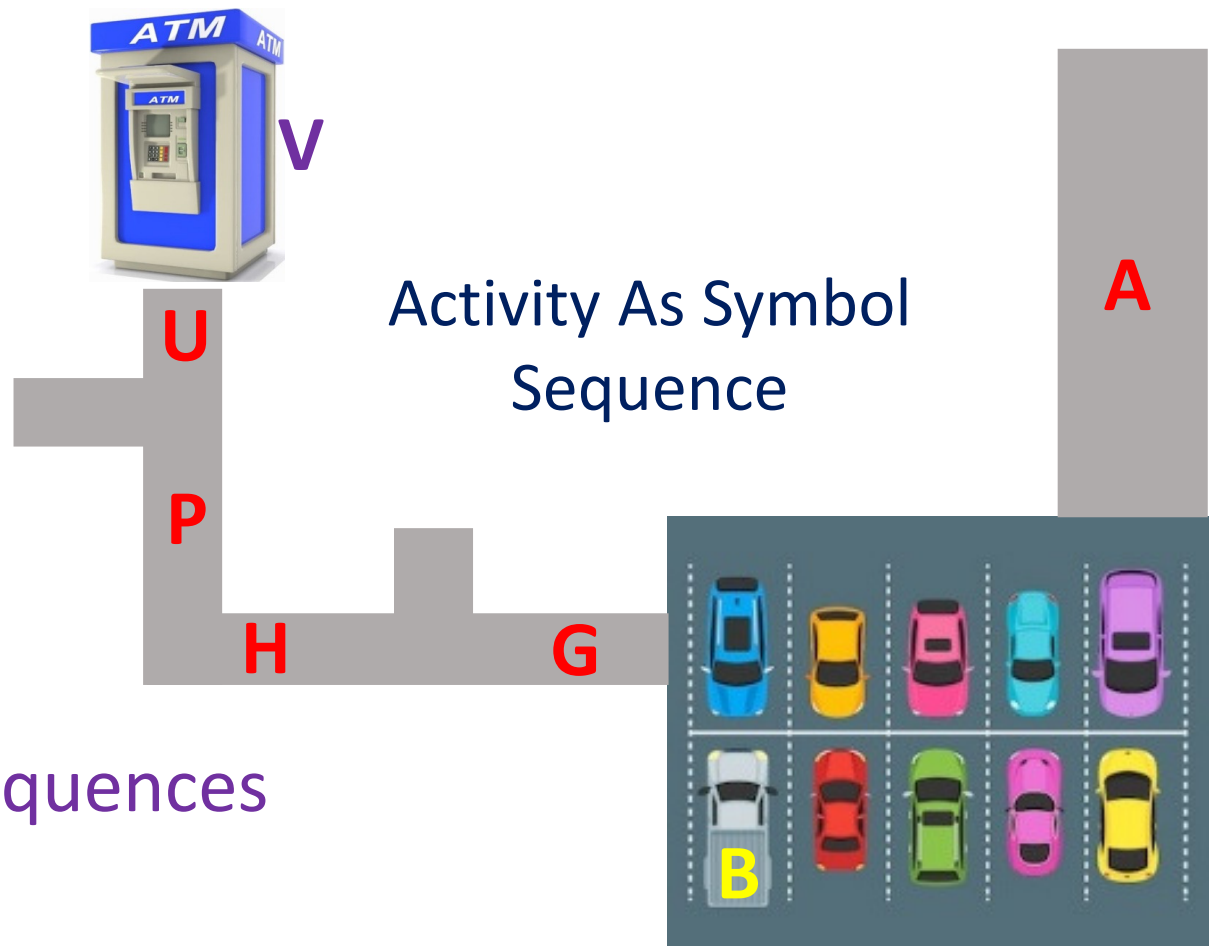


Monitoring a Shopping Area


BANK
ATM
ATM
Store
DISCOUNT
OPEN
BAKERY
OPEN
SHOP
OPEN
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W

Visiting the ATM





Activity As Symbol Sequence

Task: Modeling Symbol Sequences Generated by Activities

Time Series of Vectors

$$S_i = \left\{ x_t^{(i)}; t = 1, \ldots T_i \right\}; \mathrm{i} = 1, \ldots n$$

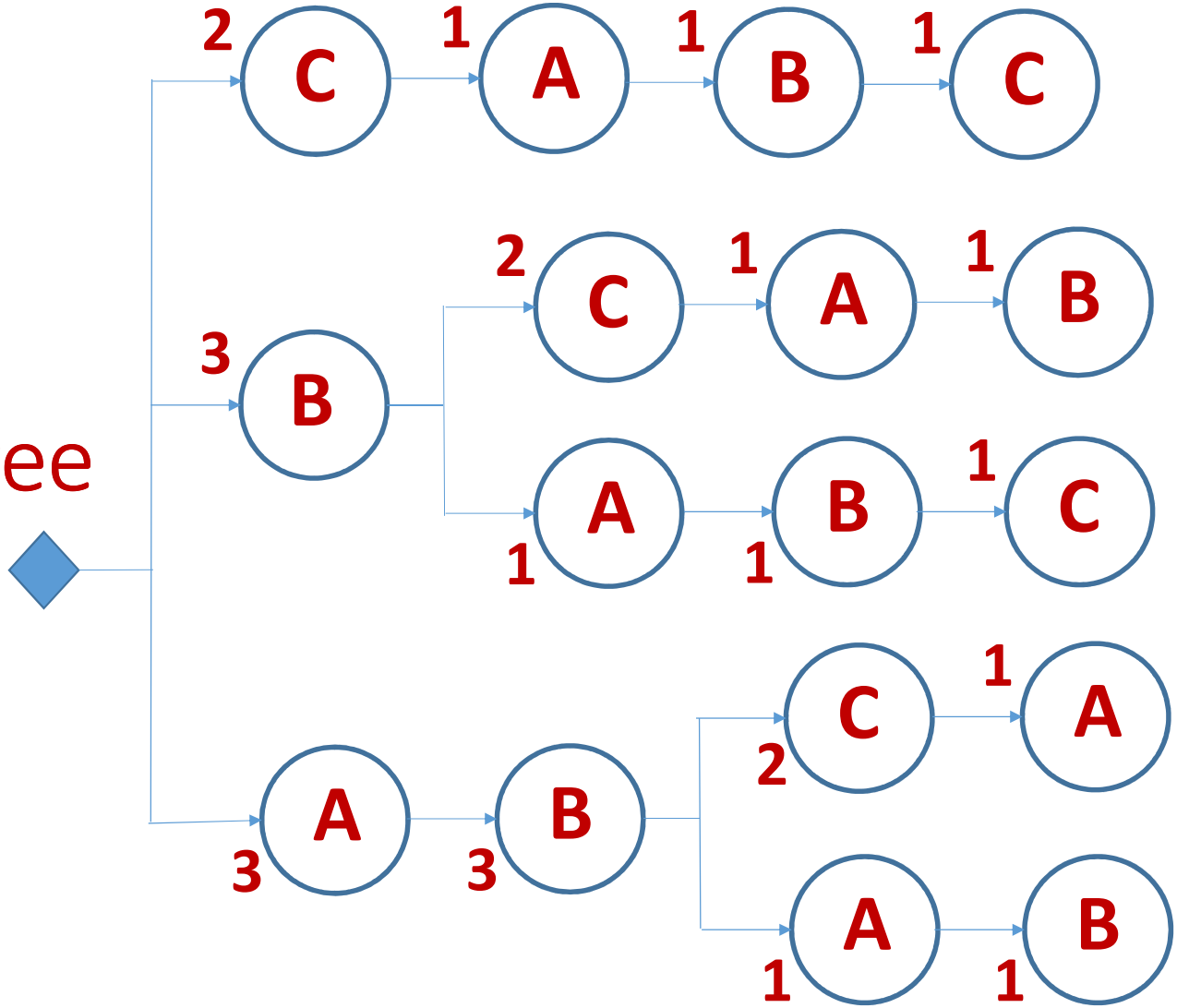
Time Series Datasets $\boldsymbol{S}_i$ are of Different Duration T_i

Task-1: Learning Model $\boldsymbol{M}_i$ for Time Series $\boldsymbol{S}_i$

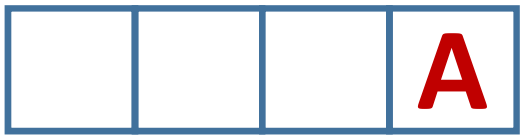
Task-2: Grouping Models $\boldsymbol{M}_i$ using Clustering Algorithms

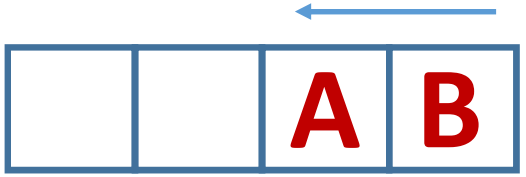
Learning a Suffix Tree

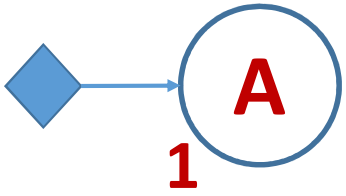
A B A B C D B D

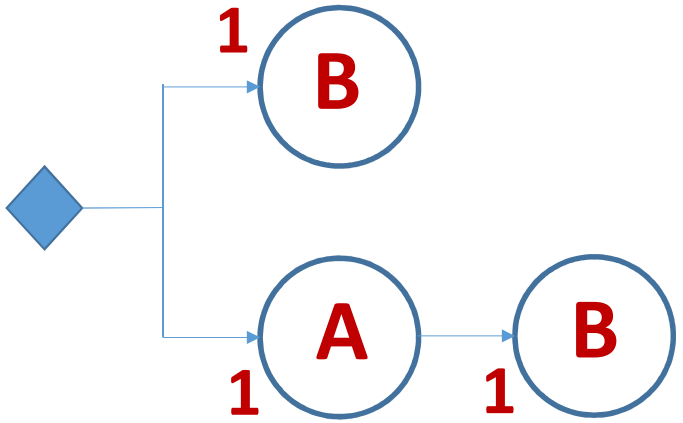


A B A B C D B D

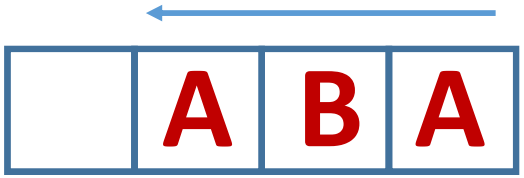


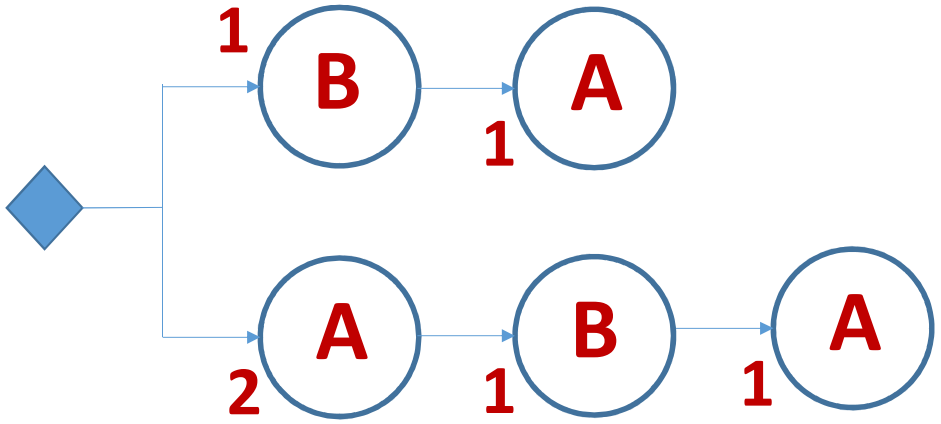




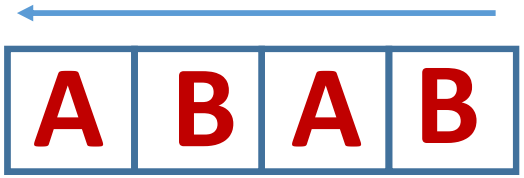


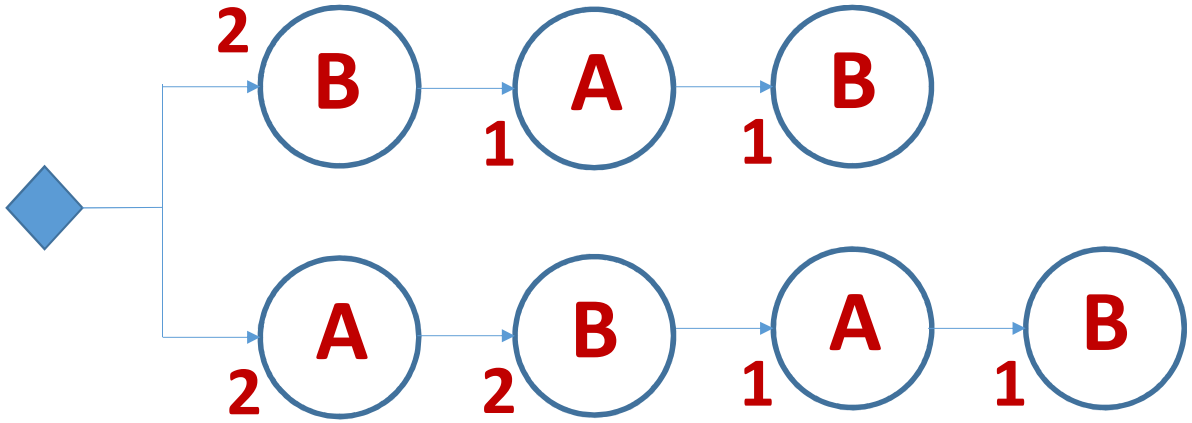
A B A B C D B D



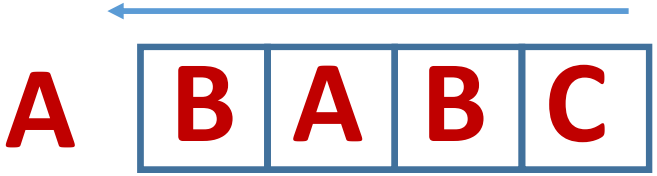


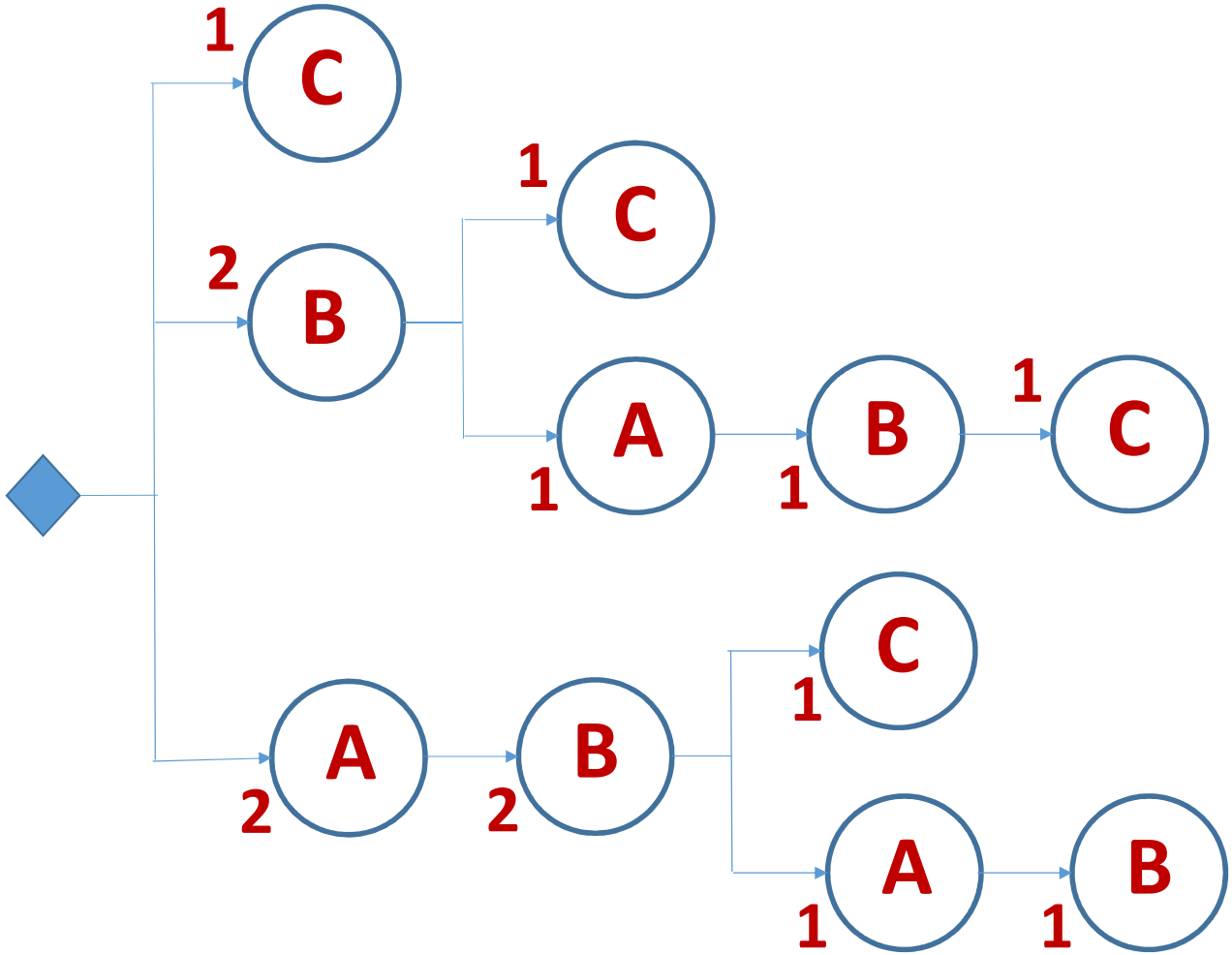
A B A B C D B D



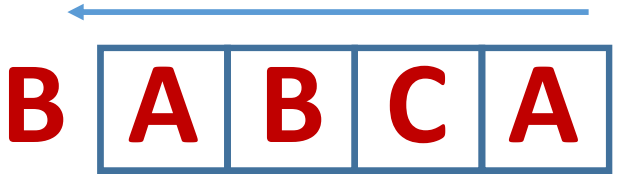


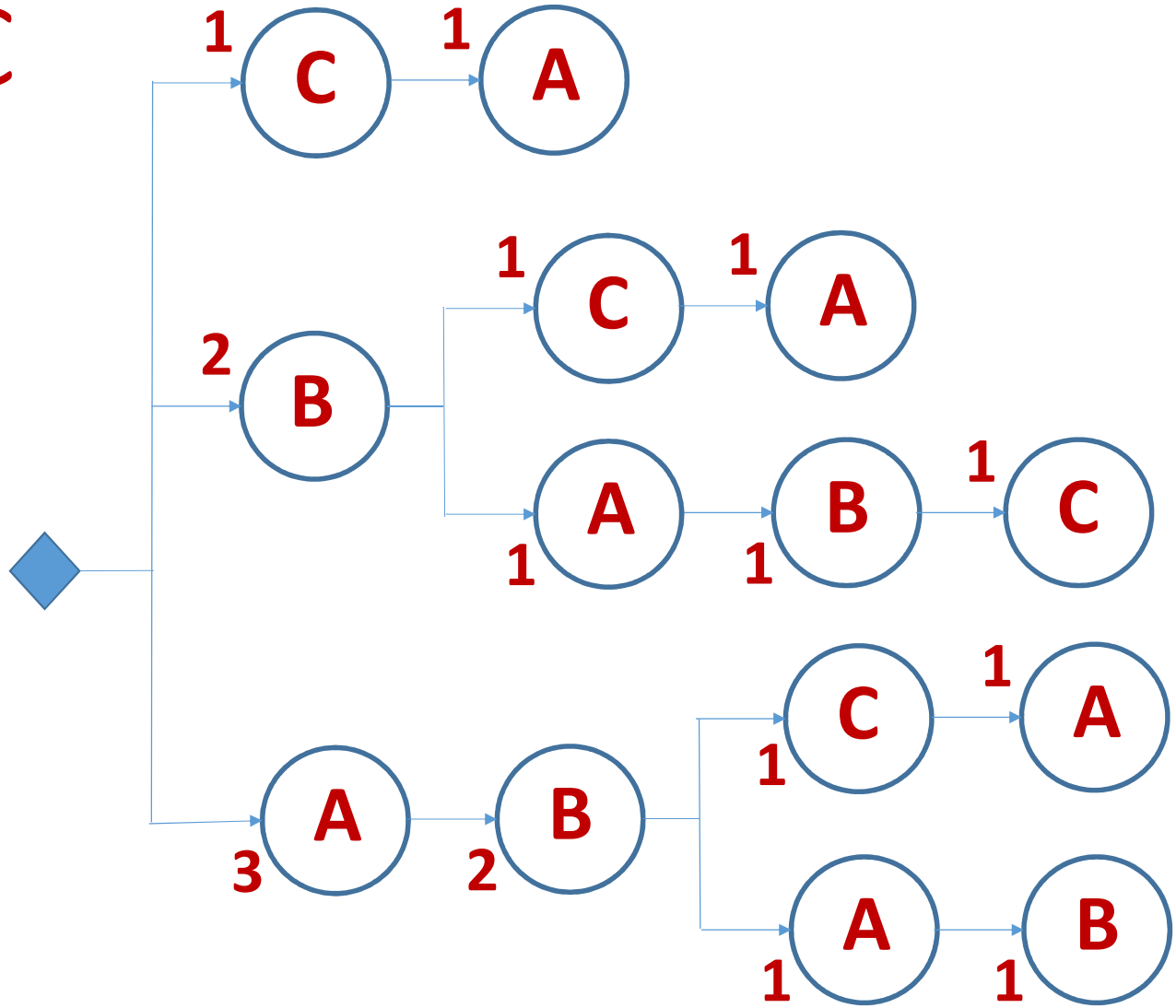
A B A B C A B C



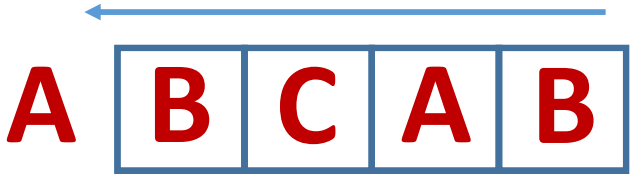


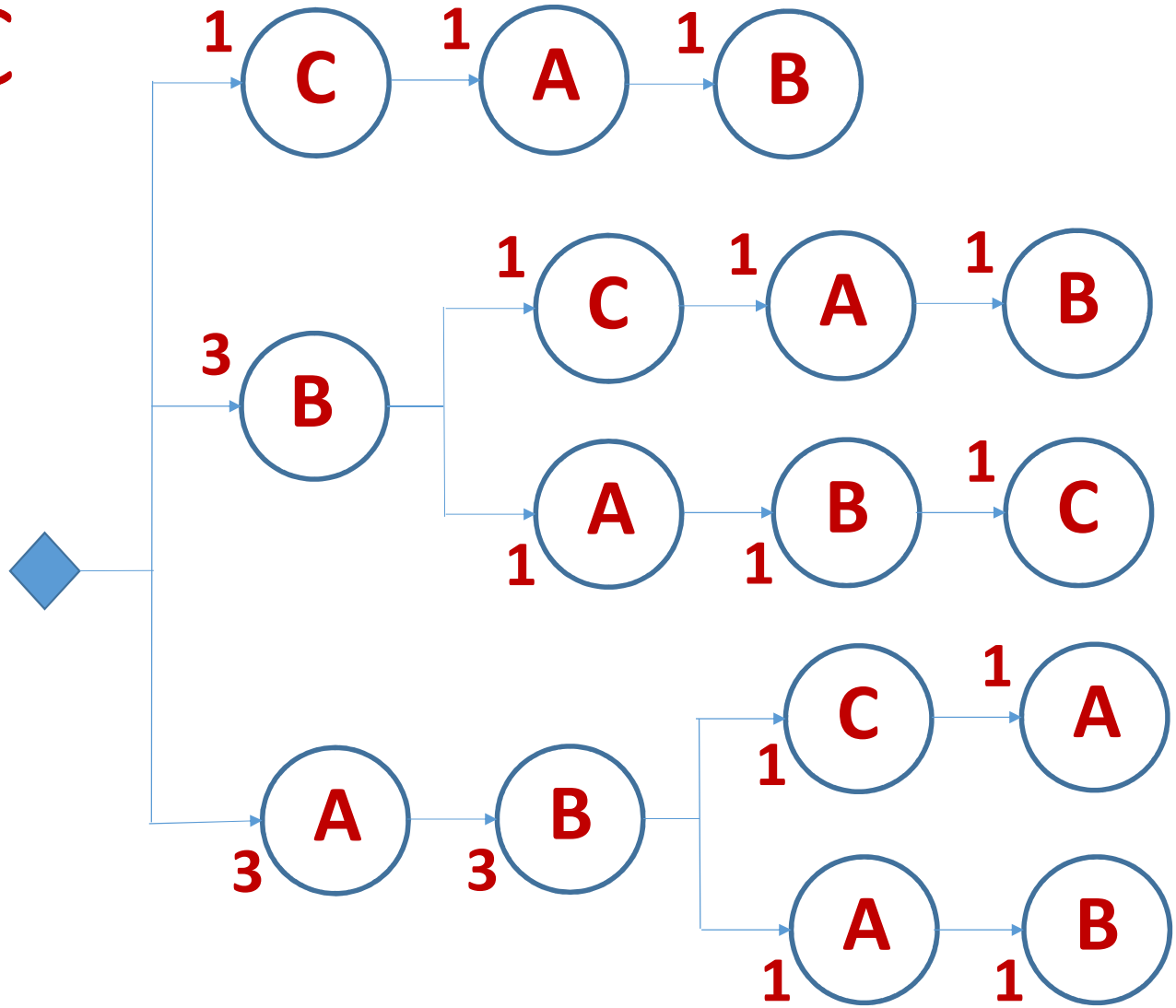
A B A B C A B C



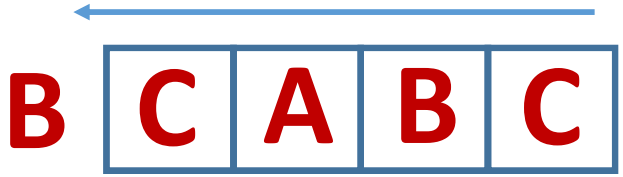


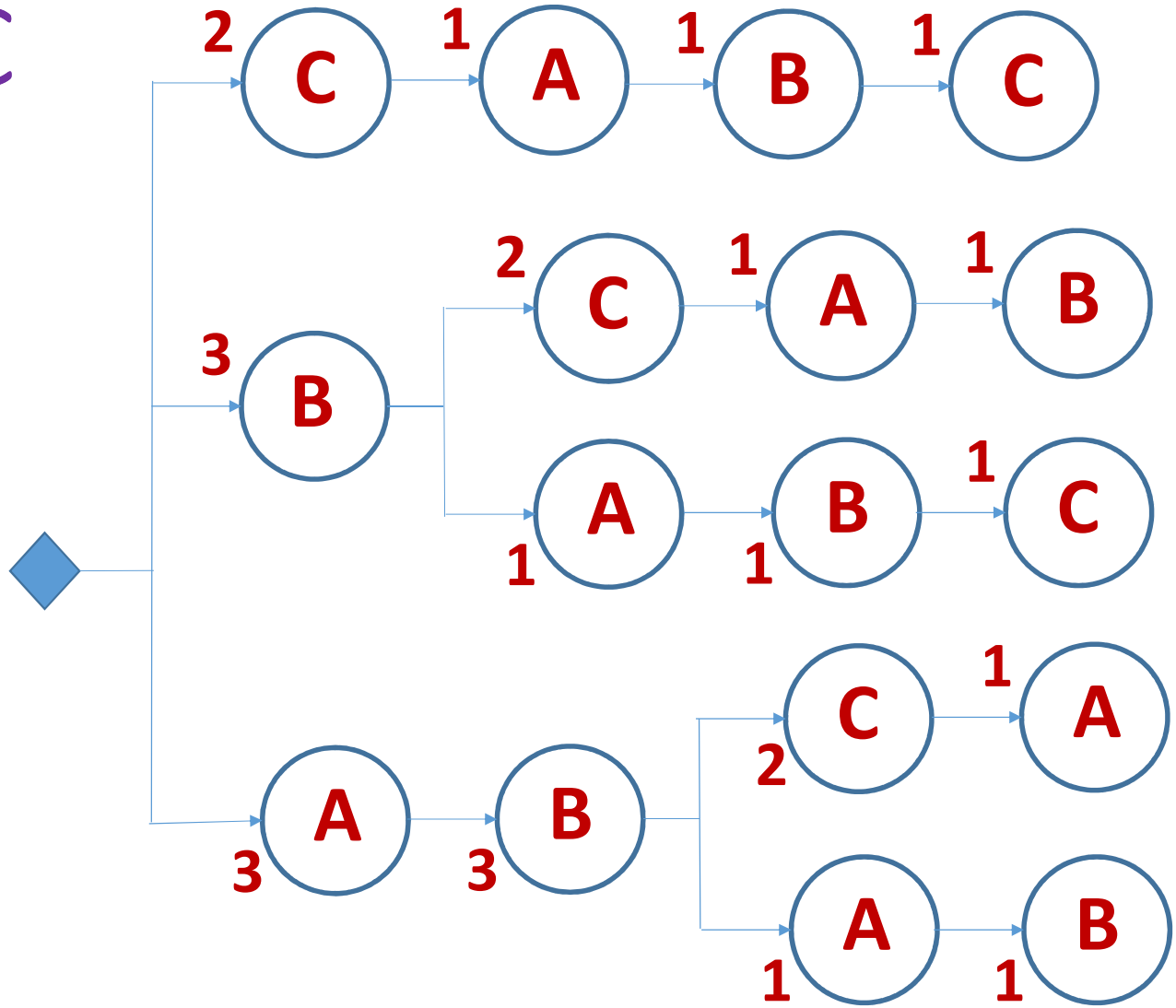
A B A B C A B C

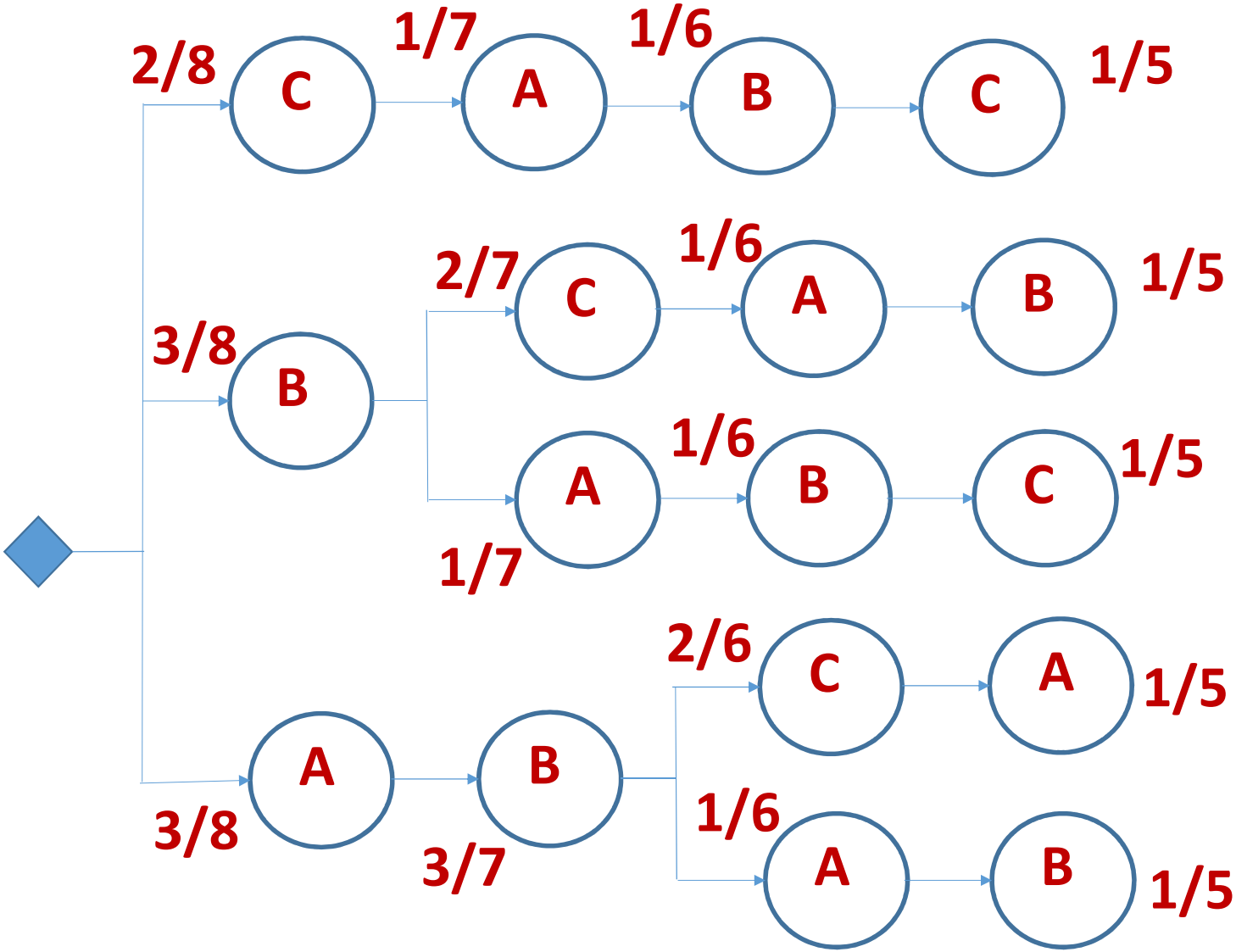




A B A B C A B C

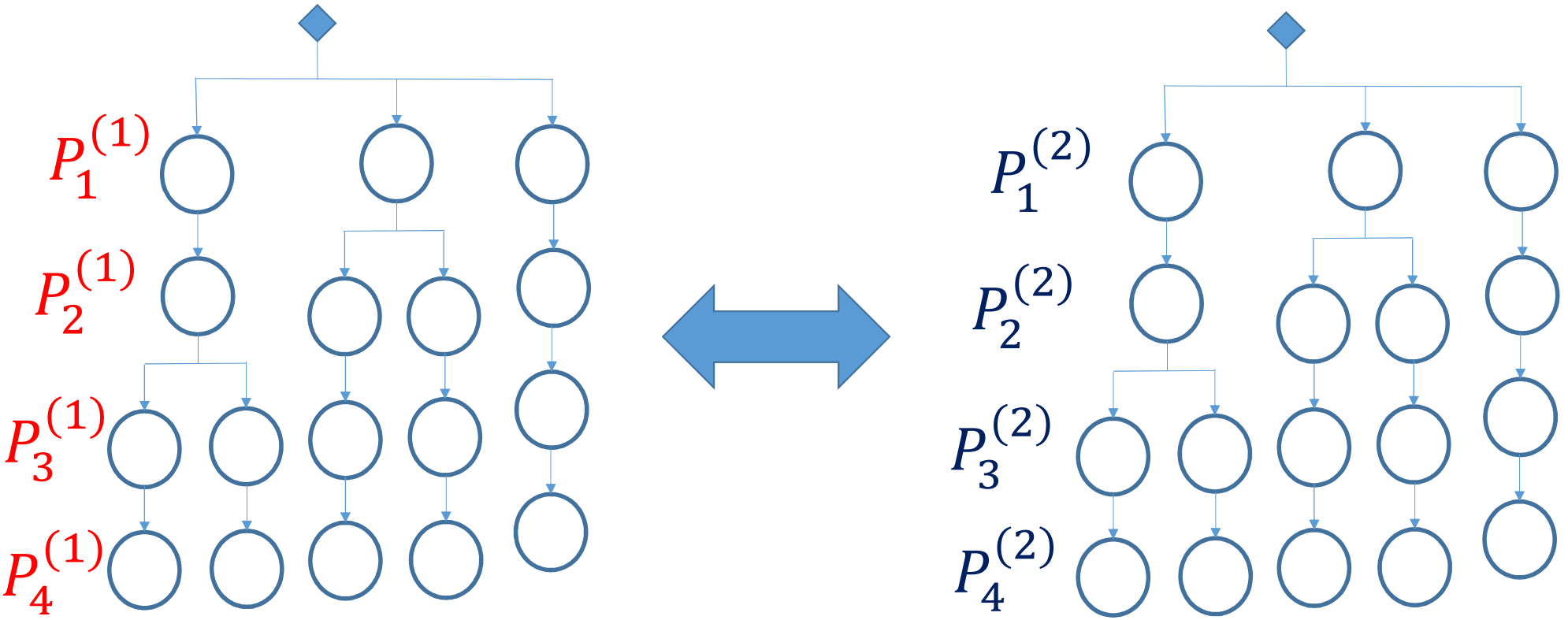






Distribution of Symbols

Similarity between Two Suffix Trees



Similarity between Two Suffix Trees

$$\mu_s(T_1, T_2) = \frac{\sum_{d=1}^{D} \omega_d BC\left(P_d^{(1)}, P_d^{(2)}\right)}{\sum_{d=1}^{D} \omega_d}$$

Handling Continuous Data

$$S_i = \left\{ x_t^{(i)}; t = 1, \dots T_i \right\}; i = 1, \dots n$$

$$\left\{ x_t^{(i)}; t = 1, \dots T_i; i = 1, \dots n \right\}$$

K-Means Clustering

Symbol Vocabulary $\{\theta_j; j = 1, \dots K\}$

Handling Continuous Data

$$S_i = \left\{ x_t^{(i)}; t = 1, \dots T_i \right\}; i = 1, \dots n$$

$$S_i \equiv \left\{ \theta_t^{(i)}; t = 1, \dots T_i \right\}; i = 1, \dots n$$

Symbol Vocabulary

$$\{\theta_j; j = 1, \dots K\}$$

Suffix Trees: Applications

- Variable Length Time Sequence Modeling
- Sequence Pattern Mining
- Sequence Grouping
- Anomalous Event Detection

Summary

- Suffix Tree Learning
- Applications in Sequence Mining
- Anomalous Event Detection
- Variable Length Sequence Grouping


भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Thank You