Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering
Department of Computer Science

Process Analytics

Prof. Dr. Agnes Koschmider
Dominik Janssen

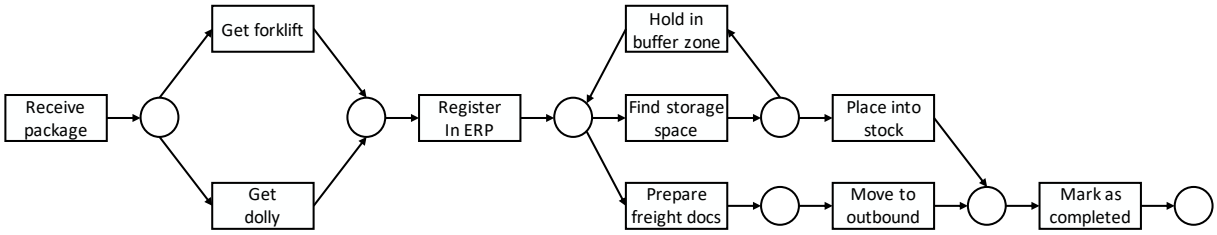
Advanced Process Mining

Summer term 2020

Exercise sheet 9

Event Log Clustering

Exercise 1: Trace Clustering



a) Cluster the process model above into

   i) 2 groups
   ii) 3 groups
   iii) 4 groups

b) What is the ideal number of clusters for this process model?

Exercise 2: Distance Measures

a) Calculate the similarity between the following traces:

$$j = \langle A, B, C, D, E, F \rangle$$
$$k = \langle A, B, A, C, D, E, F, F \rangle$$

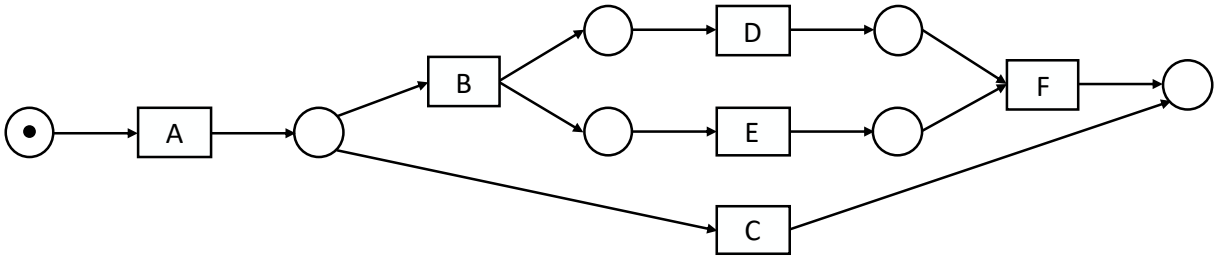
using the

   i) Euclidean distance
   ii) Hamming distance

b) In addition calculate the euclidean distance between the trace $m = \langle F, E, D, C, B, A \rangle$ and the trace j.
What is the drawback of this distance measure?

Exercise 3: Process Trees

Derive a process tree from the following Petri net:



Exercise 4: True or False

a) The k-Means clustering algorithm finds clusters of the size k.

b) The advantage of the Levenshtein distance over the Hamming distance is that it can be applied to vectors of different lengths.