Data Mining 2: Frequent Pattern Mining Summer 2018
Christian Borgelt from 16.07.2018

Exercise Sheet 11

Exercise 40 Searching for Frequent Subgraphs

a) Why do we exclude the removal of proper bridges when assigning unique parents?

b) How many different spanning trees does the graph / molecule shown on the right possess? Justify your answer!

c) Why is the number of possible code words usually (much) larger than the number of spanning trees? What additional choices does one have?

d) Why does a coding scheme based on spanning trees in which edges closing cycles are listed after the spanning tree edges ensure that we can always the last edge? (In other words: why can we spare ourselves the check that this edge is not a proper bridge? What must the last edge rather be?)

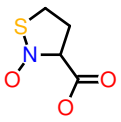
e) How many different (extended) adjacency matrices does the graph/molecule in part b) of this exercise possess?

Exercise 41 Canonical Code Words

a) Check whether the code word

`S 10-N 21-O 31-C 43-C 54-O 64=O 73-C 87-C 80-C`

is the canonical code word, based on a depth-first search spanning tree, for this graph/molecule:



Use the order `S` $\prec$ `N` $\prec$ `O` $\prec$ `C` for the atoms and the order `-` $\prec$ `=` for the bonds!

b) Check whether the code word

`S 0-N1 0-C2 1-O3 1-C4 2-C5 4-C5 4-C6 6-O7 6=O8`

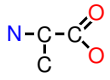
is the canonical code word, based on a breadth-first search spanning tree, for the graph/molecule of part a)! Use the same orders as in part a)!

Exercise 42 Canonical Code Words

a) What are rightmost path extensions? What are maximum source extensions?

b) Do rightmost path/maximum source extensions always yield canonical code words?

c) Find the canonical code word, based on a depth-first search spanning tree, for cyclin, that is, for the molecule:



Use the order `N` $\prec$ `O` $\prec$ `C` for the atoms and the order `-` $\prec$ `=` for the bonds!

d) Find the canonical code word, based on a breadth-first search spanning tree, for cyclin (see part c)! Use the same orders as in part c!

Exercise 43 Extendable Vertices

a) Which vertices/atoms of cyclin are extendable, based on the canonical code word from Exercise 42c), by rightmost path extensions?

b) Which vertices/atoms of cyclin are extendable, based on the canonical code word from Exercise 42d), by maximum source extensions?

Additional Exercise Perfect Extensions

a) Perfect extensions of item sets are defined by a very simple criterion.
Why is the same criterion necessary, but not sufficient for graphs?
What criterion is needed instead?

b) Why do rings cause problems to perfect extensions of graphs?
With what additional conditions can these problems be handled?
Are these conditions necessary?

c) Why is it easy to cut the search tree branches "to the right" of a perfect extension, but difficult to cut those "to the left" of a perfect extension?

d) Find the perfect extensions of the fragment `C:C` and of the fragment `O-C` in the graph database that consists of the following three molecules: