CSE 881: Data Mining (Fall 2016) Homework 7

Due date: Nov 8, 2016

1. (a) List all the 3-subsequences contained in the following data sequence:

$$< \{p, q\}\ \{r\}\ \{p, q\} >,$$

assuming no timing constraints.

(b) List all the 3-element subsequences contained in the data sequence given in part (a).

(c) List all the candidate 4-sequences produced by the candidate generation step of the GSP algorithm from the following frequent 3-sequences:

$< \{p, q, r\} >,\ \ < \{p, q\}\{s\} >,\ \ < \{p\}\{s, p\} >, < \{p\}\{p, q\} >,$
$< \{p, r, s\} >,\ \ < \{r\}\{s\}\{s\} >,\ \ < \{q, r\}\{s\} >, < \{p, r\}\{s\} >,$
$< \{q\}\{s\}\{p\} >,\ \ < \{q, r\}\{s\} >,\ \ < \{q\}\{r, s\} >,\ \ < \{r, s\}\{s\} > .$

(d) Based on your answer in part (c), list the candidate 4-sequences that survived the candidate pruning step of the GSP algorithm.

2. Consider the following data sequence s:

Timestamp	Element
10	{p, q, r, s}
20	{p, q, t}
40	{q, r}
50	{q, s, t}
80	{r, s, t}

where the timestamp indicates the time in which events associated with the given element were observed.

(a) State whether each following sequential pattern w is a contiguous subsequence of the data sequence s:

- $w =< \{p\}\{q\}\{r\}\{s\}\{t\} >$
- $w =< \{p\}\{p\}\{q\}\{q\} >$
- $w =< \{p\}\{s\}\{t\} >$
- $w =< \{p, r\}\{q, r\}\{q, s\} >$

(b) State whether the following sequential pattern is contained in (i.e., supported by) the data sequence above. Use the following time constraints:

mingap = 0, maxgap = 35, window size = 15, maxspan = 65.

- $w =< \{p, q, r, s, t\} >$
- $w =< \{q, r, s, t\}\{q, r\}\{q, s\} >$

- $w =< \{r, s\}\{r, s\}\{r, s\} >$
- $w =< \{p, q, r\}\{q, r, s\} >$

3. For each question below, draw all the candidate 5-subgraphs generated from joining a pair of frequent 4-subgraphs shown in Figure 1 using the method described in the lecture. Assume the edge-growing method is used to expand the subgraphs. Note that we focus on connected subgraphs only with no self-loops and no multiple edges (i.e., there cannot be more than one edge incident on the same pair of vertices). Indicate what are the cores (i.e., the common frequent 3-subgraphs) between the two frequent subgraphs that were joined. For example, the vertices that are part of the core can be shaded (colored) while those that are not part of the core can be left unshaded (no color). Answer no candidates if the graphs to be joined have no common core.

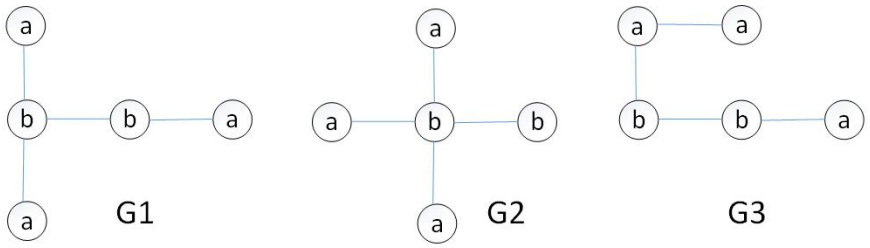

Figure 1: Frequent 4-subgraphs.

(a) Draw all the candidate subgraphs after joining G1 with itself.

(b) Draw all the candidate subgraphs after joining G2 with itself.

(c) Draw all the candidate subgraphs after joining G3 with itself.

(d) Draw all the candidate subgraphs after joining G1 with G2.

(e) Draw all the candidate subgraphs after joining G1 with G3.

(f) Draw all the candidate subgraphs after joining G2 with G3.