Group Work: TITANIC

Group 1

Question and answer 1: *How can we compute the closure of an attribute set using only the support values?* – $X'' = X \cup \{m \in M \setminus X \mid \operatorname{supp}(X) = \operatorname{supp}(X \cup \{m\})\}$

Given the slides (and the original TITANIC paper), try to answer the following questions:

1. In which lines of which algorithm (TITANIC, TITANIC-GEN, or CLOSURE) is the answer implemented?
2. What is done in line 2 of that algorithm and why does that work?
3. Given the set $\mathcal{K}_{k-1} = \mathcal{K}_4 = \{\{1,3,4,6\},\{1,3,4,7\}\}$ explain how $\mathcal{C}$ looks after line 1 of algorithm TITANIC-GEN.

Group 2

Question and answer 2: *How can we compute the closure system such that we need to compute as few closures as possible?* – We compute only the closures of the minimal generators.

Given the slides (and the original TITANIC paper), try to answer the following questions:

1. In which lines of which algorithm is the answer implemented? (hint: read the definition of *minimal generator*, i.e., key set, on slide 21 again)
2. Why is it sufficient to compare against the predecessor support as it is computed in line 5 of algorithm TITANIC-GEN?
3. Line 4 of algorithm TITANIC-GEN is also related to this answer – why?

Group 3

Question and answer 3: *How can we derive as many support values as possible from already known support values?* – If X is not a minimal generator, then $\operatorname{supp}(X) = \min\{\operatorname{supp}(K) \mid K \text{ is minimal generator}, K \subseteq X\}$.

Given the slides (and the original TITANIC paper), try to answer the following questions:

1. In which lines of which algorithm is the answer implemented?
2. What special property has the input of that algorithm?
3. How can we be sure in that line that $X \cup \{m\}$ is *not* a minimal generator?

Additional Remarks for TITANIC

TITANIC

- **line 9**: implements **answer 2**, pruning all non-key sets

TITANIC-GEN

- **line 1**: choose pairs of sets from $\mathcal{K}_{k-1}$ that have the same $k-2$ first elements, e.g., for $k=5$:

```
         k-2k-1
i = 1  2  3  4
A = 1, 3, 4, 6
B = 1, 3, 4, 7
    =  =  =  !=
```

- **line 4**: a la Apriori: prune sets with non-key/non-frequent subsets (remember: the minimal generators (= key sets) form an order ideal, i.e., if a set is *not* a key set, then none of its supersets is either)
- **line 5:** ensures that $X.p\_s$ contains the *minimum support* of all $k-1$ subsets of X *(since the support is monotonically decreasing with increasing set size, it is sufficient to compare against the minimum support of the lower covers - their support can only be larger or equal to the support of X)* - this is required for further non-key set pruning in line 9 of TITANIC (implementing **answer 2**)

CLOSURE(X)

- becomes only key sets as input because it works on sets from the previous iteration!
- **line 2**: optimization: the closures of the subsets of X have already been computed and must be contained in X'' (since $A \subseteq B \Rightarrow A'' \subseteq B''$)
- **lines 3-7**: implement **answer 1**
- **lines 5**: implements **answer 3** (if $X \cup \{m\}$ is not in $\mathcal{C}$, it can't be a minimal generator, since otherwise we would not have removed it in TITANIC-GEN (and it can't have more elements than the candidates))