Feedback — Week 3 Quiz

Help Center

Thank you. Your submission for this quiz was received.

You submitted this quiz on **Fri 27 Feb 2015 9:42 PM PST**. You got a score of **8.00** out of **9.00**.

Question 1

Given a sequence database, as shown in Table 2, with support threshold min-sup = 3, which of the following sequences are frequent?

SID	Sequence
1	$\langle a(bd)(aef)(bc)\rangle$
2	$\langle (cf)(abe)(bd)d\rangle$
3	$\langle (def)(abcde)(cde)\rangle$
4	$\langle a(abe)cd(ec)\rangle$

Table 2: Sequence database.

Your Answer		Score	Explanation
<(ae)c>	✔	0.25	<(ae)c> is a subsequence of Sequences 1, 3, and 4, so its support is 3.
<(bd)b>	✔	0.25	<(bd)b> is a subsequence of Sequence 1, so its support is 1.
<abd>	✔	0.25	<abd> is a subsequence of Sequences 2 and 4, so its support is 2.
<f(ab)>	✔	0.25	<f(ab)> is a subsequence of Sequences 2 and 3, so that the support is 2.
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "<(ae)c>".

Question 2

Given a sequence database, as shown in Table 5, and support threshold min-sup = 4, use Generalized Sequential Patterns (GSP) to find the frequent sequential patterns. After scanning the database once, how many length-2 candidate sequences will be generated after Apriori pruning? How many length-2 candidate sequences will be generated if not using Apriori pruning?

SID	Sequence
1	$\langle a(bd)(aef)(bc)\rangle$
2	$\langle (ef)(abe)(bd)d\rangle$
3	$\langle (def)(abcde)(cde)\rangle$
4	$\langle a(abe)cd(ec)\rangle$

Table 5: Sequence database.

Your Answer		Score	Explanation
None of the other options are correct.			
22; 72			
22; 51	✔	1.00	
32; 51			
32; 72			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "22; 51".

Since the min-sup = 4, after scanning the database once we have the following 4 frequent items (length-1 sequential patterns): a, b, d, e. However, we have 6 items (length-1 sequential patterns) in total, namely a, b, c, d, e, f. Therefore, the number of length 2 candidates is 4 * 4 + 4 * 3 / 2 = 22; Without Apriori pruning, the number of length 2 candidates is 6 * 6 + 6 * 5 / 2 = 51.

Question 3

Given a sequence database, as shown in Table 9, and support threshold min-sup = 4, use Generalized Sequential Patterns (GSP) to find the frequent sequential patterns. What is the minimum number of times we need to scan the database in order to find all the frequent

sequential patterns?

SID	Sequence
1	$\langle af(e)(cdeh)cfg(abe)\rangle$
2	$\langle ad(bc)c(fg)(ch)\rangle$
3	$\langle bc(ad)ebf(cdfgh)\rangle$
4	$\langle ab(bd)d\rangle$

Table 9: Sequence database.

Your Answer		Score	Explanation
4			
3	✖	0.00	
1			
> 4			
2			
Total		0.00 / 1.00	

Question Explanation

The correct answer is: "2".

Since min-sup = 4, after scanning the database once we have the following three frequent items, a, b, d.
The length-2 candidate sequences are: <aa>, <ab>, <ad>, <ba>, <bb>, <bd>, <da>, <db>, <dd>, <(ab)>, <(ad)>, <(bd)>, among which the following are frequent: <ab> and <ad>
Note that we cannot construct length-3 candidate sequences, so we only need to scan the sequence database twice.

Question 4

Given a sequence database, as shown in Table 12, and min-sup = 1, which of the following does not belong to the <e>-projected database?

SID	Sequence
1	$\langle af(e)(cdeh)cfg(abe)\rangle$
2	$\langle ad(bc)c(fg)(ch)\rangle$
3	$\langle bc(ad)ebf(cdfgh)\rangle$
4	$\langle ab(bd)d(eg)(adf)gh\rangle$

Table 12: Sequence database.

Your Answer		Score	Explanation
<(_h)cfg(abe)>	✔	1.00	
<(cdeh)cfg(abe)>			
<bf(cdfgh)>			
<(_g)(adf)gh>			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "<(_h)cfg(abe)>".

The sequences in the database are projected in the <e>-projected database as follows:

- Seq 1. <(cdeh)cfg(abe)>
- Seq 2. <>
- Seq 3. <bf(cdfgh)>
- Seq 4. <(_g)(adf)gh>

By simple comparison, we have <(_h)cfg(abe)> does not belong to <d>-projected database.

Question 5

Given a sequence database, as shown in Table 14, which of the following sequential patterns are NOT closed?

SID	Sequence	count
1	$\langle ab(ac)fd\rangle$	20
2	$\langle a(af)(ab)(cf)d\rangle$	20
3	$\langle a(be)cd(af)db(ef)\rangle$	20

Table 14: Sequence database.

Your Answer		Score	Explanation
☐ <abfd>	✔	0.25	<abfd> has support 60, and none of its superset has support 60, so it is closed.
☐ <abcfd>	✔	0.25	<abcfd> has support 40, and none of its superset has support 40, so it is closed.
☑ <(ac)fd>	✔	0.25	<(ac)fd> has support 20, and its superset <ab(ac)fd) also has support 20, so it is NOT closed.

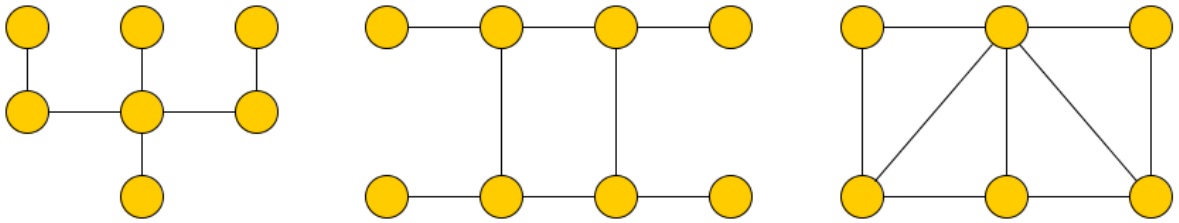
☐ <abcd>	✔	0.25	<abcd> has support 60, and none of its superset has support 60, so it is closed.
Total		1.00 / 1.00	

Question Explanation

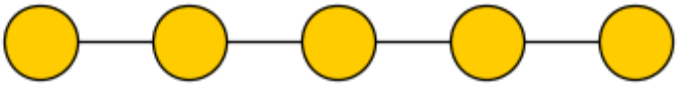
The correct answer is: "<(ac)fd>".

Question 6

In our database, we have the following three graphs:



If we set the support threshold min-sup = 3, what is the length of the longest frequent chain graph? The chain graph refers to those graphs that have nodes that are connected one-by-one. The following is an example of a length-5 chain.



Your Answer		Score	Explanation
○ 3			
○ 2			
◉ 5	✔	1.00	
○ 4			
Total		1.00 / 1.00	

Question Explanation

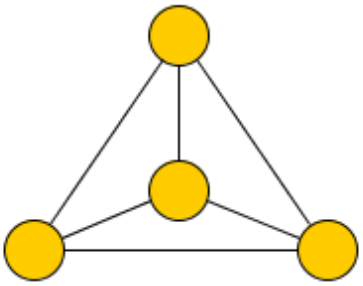
The correct answer is: "5".

The bottleneck is the first graph. We can only find a length-5 chain in the first graph (no length-

6 chain). A length-5 chain is available in the other two graphs as well.

Question 7

When we use the Apiori-based approach to find the frequent graph pattern for a candidate graph, we need to check all of its subgraphs. Given the following graph, how many distinct subgraphs with three vertices are there?



Your Answer		Score	Explanation
4			
2			
1	✔	1.00	
3			
Total		1.00 / 1.00	

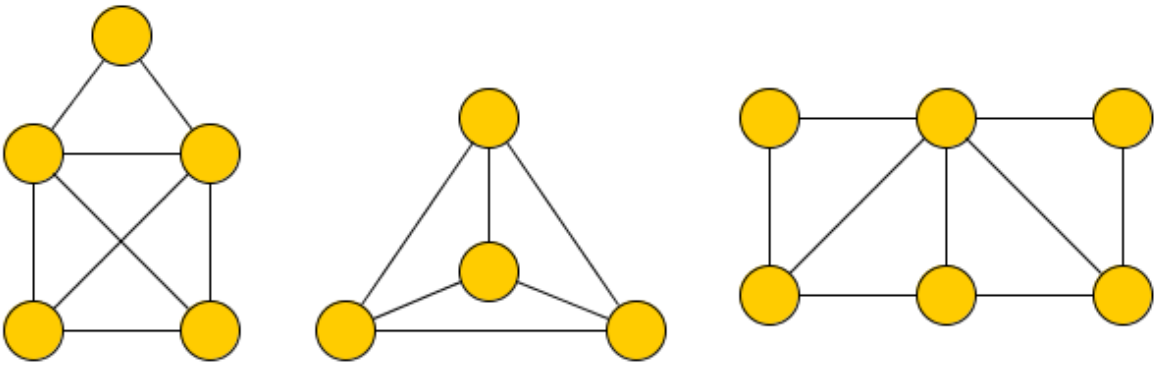
Question Explanation

The correct answer is: "1".

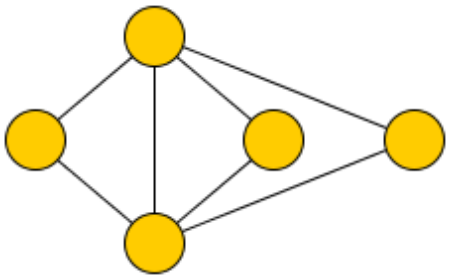
There are only four possible removes considering the connectivity. All of them lead to a triangle of three vertices. Therefore, there is only 1 distinct subgraph with three vertices.

Question 8

In our database, we have the following three graphs:



What is the support of the following graph?



Your Answer		Score	Explanation
3			
1	✔	1.00	
2			
0			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "1".

Only the first graph contains it. It needs two degree 4 nodes. Therefore, it is impossible for the last two graphs to contain it as a subgraph.

Question 9

Suppose we have learned two ranked rules as follows (the default is Type 2):

- {“ipad”, “iphone”} → Type 1
- {“kindle”, “iphone”} → Type 2

- {“ipad”} → Type 1

For the people who have {“ipad”}, which type will they be classified as by CBA algorithm?

Your Answer		Score	Explanation
◉ Type 1	✔	1.00	
○ Type 2			
Total		1.00 / 1.00	

Question Explanation

The correct answer is: "Type 1".

The first and second rules do not fit the test case, but the third rule {“ipad”} → Type 1 fits. Thus it will be classified as Type 1.