

## Quiz Rubric

### 1 Short Answer (20')

- (a) Rules with 100% confidence are obvious, not interesting, expected, everyone knows and no new knowledge about item relations is explored. If you did not explain the reason, you get 3 points.
- (b) Try different numbers for topics  $k$  and (1) calculate the log probability of topic models and pick the optimal  $k$  (2) manually observe the results and pick the optimal  $k$ . Or you can say using expert knowledge or background topics. If you only mention one reasonable way, you get 3 points.
- (c) We were not clear in our definitions in class, so we gave all answers full points (5')
- (d) All reasonable answers show possible issues with Euclidean distance are correct. Answers that do not clearly possible issues with Euclidean distance will not get credit.

### 2 Association Rules (20')

- (a) (10') - support  $\geq 2$

One-Item Set {A:6, B:7, C:6, D:2, E:2}

Two-Items Set {AB:4, AC:4, AE:2, BC:4, BD:2, BE:2}

Three-Items Set: {ABC:2, ABE:2}

(5') - confidence  $\geq 50\%$

A $\rightarrow$ B 4/6      B $\rightarrow$ A 4/7

A $\rightarrow$ C 4/6      C $\rightarrow$ A 4/6

B $\rightarrow$ C 4/7      C $\rightarrow$ B 4/6

E $\rightarrow$ A 2/2      D $\rightarrow$ B 2/2

E $\rightarrow$ B 2/2      AB $\rightarrow$ C 2/4

AC $\rightarrow$ B 2/4      BC $\rightarrow$ A 2/4

AB $\rightarrow$ E 2/4      AE $\rightarrow$ B 2/2

BE $\rightarrow$ A 2/2

- (b) (5')

Drawback: confidence  $\neq$  correlation; not interesting rules; not personalized

Solution: reasonable solution related to the drawback

### 3 Graphs (20')

This was a bit tricky, as there are several different versions of conductance. We talked about one version in class, but there are others. So we

- (a) (7')

Based on the slides:  $2/\min(8, 22-8) = 2/\min(14, 22-14) = 2/8 = 1/4$

Based on Wiki definition [https://en.wikipedia.org/wiki/Conductance\\_\(graph\)](https://en.wikipedia.org/wiki/Conductance_(graph))

$2/\min(5, 13-5) = 2/5$

- (b) (7')

Based on the slides:  $2/\min(12, 22-12) = 2/\min(10, 22-10) = 2/10 = 1/5$

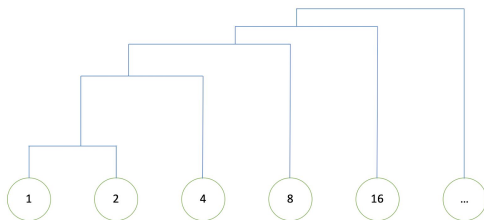
Based on the Wiki definition [https://en.wikipedia.org/wiki/Conductance\\_\(graph\)](https://en.wikipedia.org/wiki/Conductance_(graph))  
 $2/\min(6,13-6) = 1/3$

(c) (6')

Both are reasonable, depending on your perspective.

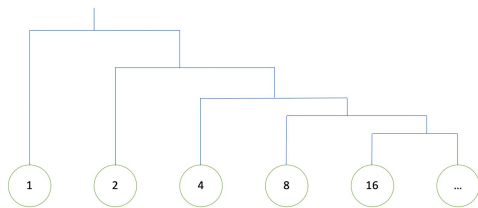
#### 4 Hierarchical Clustering (20')

(a) (6')



(b) (tree is same as (a)) (6')

(c) Yes (2') + new tree (6')



#### 5 Topic Modeling (20')

(a) (7')

$$p(word) = \frac{p(z=0)p(z=0)}{p(z=0)p(z=0)+p(z=1)p(z=1)}$$

$$p(word) = \frac{p(z=1)p(word|z=1)}{p(z=0)p(word|z=0)+p(z=1)p(word|z=1)}$$

(b) (7')

$$\log \text{likelihood} = \sum_{word} \text{count}(word) \log(p(z=0)p(z=0) + p(z=1)p(z=1))$$

(c) (6')

Any reasonable strategies to summarize topic are correct. (Such as use words with highest probabilities in each topic to represent the topic). Answers that do not clearly show how you summarize a topic will not get credit.