

THE HONG KONG POLYTECHNIC UNIVERSITY  
HONG KONG COMMUNITY COLLEGE

**Subject Title** : Introduction to Big Data Analytics      **Subject Code** : CCN3163

**Session** : Semester Two, 2017/18

**Numerical Answers**

**Question C1**

(a)

	Stat	Words						
		John	likes	apple	Peter	hates	really	orange
All Documents	IDF	1.1	0.41	0.41	0.41	1.10	1.10	1.10
Document 1	TF	0.33	0.33	0.33	0	0	0	0
	TF.IDF	0.37	0.14	0.14	0	0	0	0
Document 2	TF	0	0	0.33	0.33	0.33	0	0
	TF.IDF	0	0	0.14	0.14	0.37	0	0
Document 3	TF	0	0.25	0	0.25	0	0.25	0.25
	TF.IDF	0	0.1	0	0.1	0	0.27	0.27

(c)

$$\begin{bmatrix} 0 & 0.55 & 0.56 \\ 0.55 & 0 & 0.56 \\ 0.56 & 0.56 & 0 \end{bmatrix}$$

**Question C2**

(a)

1-itemsets	{Apple}	{Banana}	{Cabbage}	{Donut}	{Egg}
Support	3/5=0.6	2/5=0.4	1/5=0.2	2/5=0.4	4/5=0.8

(b) Support=0.2, Confidence=0.25, Lift=0.625  
(c) Support=0.2, Confidence=0.5, Lift=2.5

**Question C3**

- (a)(ii) 0.4  
(a)(iii) 3  
(b)(i)  $d(A, B) = 1, d(A, C)=2, d(A, D) = 2, d(A, E) = 1,$   
 $d(B, C) = 1, d(B, D) = 1, d(B, E) = 1,$   
 $d(C, D) = 2, d(C, E) = 2,$   
 $d(D, E) = 1$

(b)(ii)

Vertex	Normalized closeness centrality
Amy	$\frac{4}{(1 + 2 + 2 + 1)} = 0.67$
Bill	$\frac{4}{(1 + 1 + 1 + 1)} = 1$

Candy	$\frac{4}{(2 + 1 + 2 + 2)} = 0.57$
Danny	$\frac{4}{(2 + 1 + 2 + 1)} = 0.67$
Eric	$\frac{4}{(1 + 1 + 2 + 1)} = 0.8$

**Question C4**

(a)(ii)

Iteration	$PR_1$	$PR_2$	$PR_3$	$PR_4$
0 (Initial values)	0.25	0.25	0.25	0.25
1	0.25	0.11	0.11	0.53
2	0.49	0.11	0.11	0.30
3	0.29	0.18	0.18	0.36

(b)(ii) 3.42

(b)(iii) 0.0175