

## בוחר במודלים במסחר אלקטרוני (096211)

סמסטר אביב 2004

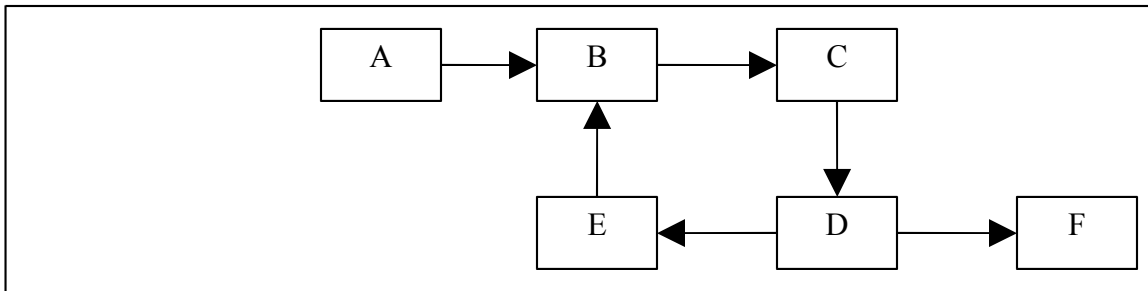
משך הבחינה: שעותיים

יש להגיש את טופס הבחינה!

שימוש בחומר עזר אסור מלבד מחשב כיס!

### שאלה 1

נתון mini-web המורכב מדפים A, B, C, D, E, F באופן הבא:



1. מהו הדף בעל ה-PageRank הגבוה ביותר (כאשר  $\text{damping factor} = 85\%$ )? נמק בפירוט (הראה את החישובים) והסבר מדוע התקבלה תוצאה זו.
2. הבעלים של דף F מעוניינים להגדיל את ה-PageRank של דף זה על-ידי הוספת קישור אחד בלבד מדף כלשהו לדף כלשהו. מהו הקישור אותו היית ממליץ להוסיף כך שהגידול בערך ה-PageRank של דף F יהיה מקסימאלי? נמק תשובתך והראה מהו ערך ה-PageRank של דף F לאחר הוספת הקישור שהצעת.

### שאלה 2

נתונה קבוצת הנקודות הבאות על ציר המספרים: {6, 12, 18, 24, 30, 42, 48}.

1. מהם ה-2 clusters שיוצרו לאחר הרצת אלגוריתם hierarchical clustering? פרט את כל צעדי האלגוריתם. (הנח מטריקה אוקלידית סטנדרטית)
2. מהם ה-2 clusters שיוצרו לאחר הרצת אלגוריתם k-means? פרט את כל צעדי האלגוריתם וציין מהם ה-centroids בכל אחד מהצעדים. (הנח מטריקה אוקלידית סטנדרטית)
3. תן דוגמא למקרה בו קיימים n clusters אולם אחד מאלגוריתמי ה-clustering שלמדנו (hierarchical clustering) יצליח לבנות n-1 clusters בלבד. נמק מדוע הדוגמא שהבאת מקיימת את התנאי הנ"ל.  
הערה: לצורך הבאת הדוגמא ניתן לבחור  $n > 2$  כלשהו.

נתון אוסף טרנזקציות market-basket הבא:

פריטים שניקנו	מספר טרנזקציה	מספר לקוח
A, D, E	0001	1
A, B, C, E	0024	1
A, B, D, E	0012	2
A, C, D, E	0031	2
B, C, E	0015	3
B, D, E	0022	3
C, D	0029	4
A, B, C	0040	4
A, D, E	0033	5
A, B, E	0038	5

1. חשב את ה- support עבור המוצרים:  $\{E\}$ ,  $\{B, D\}$ ,  $\{B, D, E\}$ .

2. חשב את רמת הביטחון של ה- association rules הבאים:  $BD \rightarrow E$ ,  $E \rightarrow BD$ . הסבר את התוצאות.

3. בנה מטריצת market-basket עפ"י הנתונים בטבלה הנ"ל לפי מספרי הלקוחות כאשר כל מוצר הוא משתנה בינארי (ערכו 1 אם הוא מופיע בלפחות טרנזקציה אחת של הלקוח ו-0 אחרת). פתור את סעיפים א' ו-ב' על מטריצה זו.

4. הנח כי  $s_1$  הוא ה- support של association rule  $r$  עפ"י הטבלה הנתונה. הנח כי  $s_2$  הוא ה- support של

association rule  $r$  עפ"י המטריצה שבנית בסעיף ג'.

האם ישנו קשר ( $=<>$ ) בין  $s_1$  ו-  $s_2$ ? הוכח תשובתך!

#### שאלה 4

שרון מייצרת זוג מפתחות פומבי ופרטי על מנת לתקשר עם יואב. היא שולחת ליואב את המפתח הפומבי שלה וכמו כן מקבלת מיואב את המפתח הפומבי שלו. בהינתן כי שרון שולחת הודעות ליואב כך שהפעולה היחידה שהיא מבצעת היא חתימה עליהן.

1. ציין מהם היתרונות והחסרונות של צורת עבודה זו.

2. לגבי כל אחת מאפשרויות הבאות הסבר האם ניתן לפתור בעזרתה את החסרונות שציינת. נמק תשובתך!

Blind signature.1

Off-line escrow service.2

3. הצפנת כל ההודעות ע"י המפתח הפומבי של יואב

Micro payments.4

לכל אחת מהטענות הבאות קבע האם היא נכונה/לא נכונה. נמק!

הנתונים הבאים מתייחסים לטענות 1 ו-2 בלבד:

נגדיר new Jaccard measure להיות:  $\frac{d}{b+c+d}$ .

נגדיר אלגוריתם new Min hashing:

• בצע פרמוטציה אקראית על השורות.

•  $h(c_i)$  = אינדקס השורה הראשונה בה מופיע 0 בעמודה  $c_i$ .

טענה 1:  $P[h(c_i)=h(c_j)] = Sim_{newJaccard}(c_i, c_j) \quad i \neq j$ .

טענה 2: שימוש ב- new Jaccard measure שקול לשימוש ב- Jaccard measure המקורי.

טענה 3: באלגוריתם Hubs & Authorities: האלגוריתם האיטרטיבי של Kleinberg שקול לחלוטין לאלגוריתם

החישובי (המשתמש במטריצת הסמיכויות של ה-web).

בהצלחה!