

# תרגיל בית 1 - ניתוח רשתות

## תאריך הגשה: ראה באתר הקורס

### מטרה

מטרת התרגיל היא התנסות בשיווק באמצעות רשתות חברתיות, ובו עליכם לשווק דיסקים חדשים של מספר אמנים דרך רשת חברתית.

כקלט תקבלו רשת חברתית אמיתית של כמעט 2000 משתמשים בתחנת רדיו אינטרנטית, יחד עם רשימה (אנונימית – שמות מוחלפים במספרים) של אמנים המושמעים בתחנת רדיו זו, בנוסף, עבור כל משתמש, תקבלו את מספר הפעמים ששמע שיר של כל אחד מהאמנים.

מתוך רשימה של 10 אמנים תבחרו 6, ולכל אחד מהאמנים שבחרתם תשווקו את הדיסק שלו. לכל דיסק שאתם צריכים לשווק, עליכם לבחור 5 משתמשים. כל משתמש כזה יקבל במתנה את הדיסק החדש של האמן. כל חבריו של משתמש זה יראו שהמשתמש "קנה" את הדיסק, ויחליטו אם לקנות את הדיסק (או לא) לפי מודל אקראי, המפורט בהמשך:

### מודל התפשטות

במודל ההתפשטות קיימת הנחה של TRIADIC CLOSURE (תזכורת: אם לקדקוד יש שני קדקודים חברים, יש לצפות שבבוא הזמן יסגר משולש - תיווצר קשת בין 2 חברי הקדקוד). ברשת החברתית הנתונה, בין כל שתי נקודות זמן  $(t-1, t)$  קיים סיכוי מצטבר של  $p=0.02$  להיווצרות קשת בין שני קדקודים בעלי חבר משותף. הערה: בסיכוי מצטבר, הכוונה היא לכך שאם ל-2 לקדקודים יש  $k$  חברים משותפים הסיכוי להיווצרות קשת בין כל 2 נקודות זמן הוא:  $1 - (1-p)^k$

כאשר משתמש מסוים קונה דיסק של אמן בנקודת זמן  $t$  (כולל המשתמשים הראשוניים שמקבלים מכך את הדיסק במתנה בזמן  $t=0$ ), החברים שלו רואים זאת בנקודת זמן  $t+1$ . כעת, כל אחד מחברים אלה מקבל החלטה אם לקנות את הדיסק או לא, באופן אקראי. ההסתברות של משתמש לקנות דיסק תלויה במספר גורמים:

- מספר החברים של המשתמש -  $d$
- מספר החברים של המשתמש שקנו את הדיסק -  $b$
- יחס המשתמש אל האמן

וביחס המשתמש אל האמן, שיכול להיות:

- לא מכיר (אם המשתמש לא שמע את האמן אף פעם)
- מכיר (ואז מספר הפעמים שהמשתמש שמע שיר של האמן ידוע - נקרא לו  $a$ )

אם המשתמש לא מכיר את האמן, ההסתברות שהוא יקנה את הדיסק החדש שלו שווה למספר החברים שלו שקנו את הדיסק, חלקי המספר הכולל של החברים שלו, כלומר  $b/d$ .

אם המשתמש מכיר את האמן, ההסתברות שהוא יקנה את הדיסק החדש שלו שווה למספר החברים שלו שקנו את הדיסק, חלקי מספר החברים שלו סך הכל, ועוד מספר הפעמים שהוא שמע את האמן חלקי 2000, כלומר  $\min(1, b/d + a/2000)$ .

**לדוגמה:** עבור משתמש בעל 8 חברים, מתוכם 3 קנו דיסק חדש של אמן שהוא לא שמע אף שיר שלו, יקנה את הדיסק בהסתברות של  $3/8$ . לעומת זאת, אם אותו משתמש שמע שיר של אותו אמן 2000 פעמים, הוא יקנה את הדיסק בהסתברות של 1, ואם הוא שמע שיר של אותו אמן 100 פעמים, הוא יקנה את הדיסק בהסתברות  $17/40$ .

**שימו לב,** בכל נקודת זמן  $t$ , כל משתמש שלא קנה עדיין דיסק מסוים, מקבל החלטה האם לקנות דיסק זה ע"פ המודל ההסתברותי המתואר. הרווח שלכם על כל קניית דיסק שווה ל-1. כל משתמש יכול לקנות (או לקבל) לכל היותר דיסק אחד של אותו אמן. כלומר, אם בחרתם לקדם את האמנים 70 ו-150, לכל משתמש יכולים להיות דיסק של אמן 70 ודיסק של אמן 150, אך לא שני דיסקים של אמן 70.

בתרגיל זה תקבלו רשימה של 10 אמנים מתוכם עליכם לבחור 6. עבור כל אמן שבחרתם עליכם לבחור 5 משתמשים שיקבלו את הדיסק החדש של האמן במתנה, על מנת למקסם את תוחלת הרווח שלכם לאחר נקודת זמן  $t=6$ , בדיוק אחרי שהחלטות של  $t=6$  נלקחות.

## ציון

אנו נבצע מספר הרצות אקראיות המסמלצות את המודל. הציון שלכם בתרגיל זה ייקבע לפי הרווח הממוצע שלכם לאחר נקודת הזמן  $t=6$ .

## ממשק

כקלט, תקבלו את הקבצים הבאים:

הקובץ `user_friends.csv` מתאר את הרשת החברתית - כל שורה בקובץ מכילה זוג מזהים של משתמשים, ומשמעותה ששני משתמשים אלה הם חברים (הקשתות ממומשות בשני כיוונים). לדוגמה, אם יש קשת בין A ל-B, גם  $[A,B]$  וגם  $[B,A]$  יופיעו בקובץ.

הקובץ `user_artists.csv` מתאר כמה פעמים כל משתמש שמע שיר של כל אמן. כל שורה בקובץ מכילה שלושה שדות: `userID` - מזהה המשתמש, `artistID` - מזהה האמן, ו `weight` - כמה פעמים משתמש זה שמע שיר כלשהו של אמן זה.

בשני הקבצים, השורה הראשונה מתארת את שמות השדות, השדות מופרדים אחד מהשני ע"י פסיק.

לכל אמן שבחרתם לקדם, עליכם ליצור קובץ אחד שיכיל את מזהי המשתמשים שיקבלו את הדיסק החדש של אותו אמן בחינם. שם הקובץ עבור כל אמן יהיה `artist_artistID`, כאשר `artistID` הוא המזהה של האמן. לדוגמה אם בחרתם לקדם אמן 70, שם הקובץ עבור אותו אמן יהיה `artist_70`.

רשימת האמנים ממנה עליכם לבחור 6 אמנים היא:

70  
150  
989  
16326  
144882  
194647  
389445  
390392  
511147  
532992

יחד עם הקבצים user\_friends.csv ו-user\_artists.csv תקבלו קובץ נוסף ex1.py שבו מופיעה דוגמה לקליטת הנתונים ולפלט שעליכם לייצר.

### הוראות הגשה

עליכם להגיש **קובץ zip יחיד** (שימו לב, קובץ בפורמט ZIP בלבד - לא RAR ולא שום פורמט אחר) המכיל קובץ אחד עבור כל אחד מהאמנים שעליכם לקדם. בנוסף, עליכם להגיש תיקיה ובה התוכנה בה השתמשתם ליצירת ההמלצות (הקוד עצמו), **בנוסף בתיקיה זו עליכם לספק קובץ עם תיאור קצר של הקוד והאסטרטגיה בה בחרתם**. כל אלו חייבים להימצא בתיקיה העליונה (כלומר, שלא תהיה תיקיה בתוך קובץ ה-zip). שם קובץ ה-zip צריך להיות ID1\_ID2.zip (או ID1.zip אם מגיש יחיד) כאשר ID1 ו-ID2 הם מספרי תעודות הזהות של המגישים. רק אחד מבני הזוג צריך להגיש את התרגיל.

ההגשה ביחידים או בזוגות **בלבד**