

## רקע

חברת בגדים יוקרתיים חדשה (Praducci) מעוניינת להגדיל את ערך המותג שלה ברשת החברתית "NoseBook" ולשם כך רוצה להתפרסם. לכן, בתור סטודנטים מומחים למסחר אלקטרוני, חברה זו פנתה אליכם וביקשה שתיעצו ותסייעו לה בפרסום.

הרשת החברתית "NoseBook" מאפשרת למשתמשים להעלות סרטונים (ובפרט, סרטונים בהם מראים את הבגדים החדשים שלהם). בכל פעם שמשתמש רוכש בגד, חבריו ברשת יכולים לראות זאת לאחר מכן. לצורך ביצוע המשימה קיבלתם מסד נתונים של חברת NoseBook המתאר את הקשרים החברתיים בין המשתמשים.

## משימה

Praducci מביאה לכם 1000 דולר כדי לקנות משפיענים (Influencers) ולהביא להם בגד של החברה בחינם בזמן  $t = 0$ . כאשר משתמש מקבל/קונה בגד של Praducci, החברים שלו ברשת מחליטים לקנות בגד של החברה בהסתברות מסוימת. בנוסף, יש הסתברות מסוימת שמשתמשים ששכניהם קנו/קיבלו את המותג עוד לא ראו את מאורע הקנייה ברשת החברתית לאחר סיבוב 6, ולכן עוד לא נחשפו לחברה. המטרה שלכם היא שכמה שיותר משתמשים יכירו את המותג לאחר 6 תקופות זמן ( $t = 1, \dots, 6$ ). שימו לב ששכירת שירותיו של משפיען כרוכה בתשלום כפי שיוסבר בהמשך.

## הסתברות רכישה

- ההסתברות שמשתמש ירכוש בגד תלויה בחבריה ברשת NoseBook. את הבגדים נשלח למשפיענים בזמן  $t = 0$ , ושכניהם (חבריהם) יוכלו לראות זאת רק בנקודת הזמן הבאה, כלומר ב  $t = 1$ . בזמנים  $t = 1 \dots 6$ , כל משתמש בוחר אם לרכוש בגד של החברה לפי הסתברות ידועה מראש, באופן הבא.
- נסמן ב-  $N_t$  את מספר השכנים של המשתמש בזמן  $t$ .
  - נסמן ב-  $B_t$  את מספר השכנים של המשתמש שקנו/קיבלו את בגד של Praducci עד זמן  $t$ , **לא כולל**.

ההסתברות לרכישה עבור כל משתמש הינה  $\frac{B_t}{N_t}$ .

**הערה:** ההסתברות הנ"ל תקפה בכל נקודת זמן. במילים אחרות, אם משתמש בחר שלא לקנות בגד בנקודת זמן מסוימת, הוא עדיין יכול לעשות זאת בנקודות הזמן הבאות – עפ"י ההסתברות המתאימה.

## הסתברות חשיפה בזמן $t = 6$

ההסתברות שמשתמש  $x$  שלא קנה בגד של החברה ושקיים לו שכן שקנה בגד של החברה יהיה חשוף לחברה לאחר הסיבוב  $t = 6$  שווה ל:

$$p(x) = \frac{1}{1 + 10 \cdot e^{-\frac{B_7(x)}{2}}}$$

כאשר  $B_7(x)$  היא כמות השכנים של המשתמש שקנו בגד של Praducci עד זמן  $t = 6$  (כולל). ההסתברות שמשתמש שאף אחד משכניו לא קנה בגד של החברה היא 0.

## עלות בחירת צומת

כיום משפיענים ברשת לא מסכימים לפרסם מוצרים ללא תמורה כספית. לכן, בחירה של צומת כמשפיען שיפרסם את הבגדים של חברת Praducci כרוכה בעלות מסוימת. את המחיר **החד פעמי** שדורש כל משתמש (שעשוי להיות גם משפיען) עבור פרסום המוצר תוכלו למצוא בקובץ costs.csv. כל שורה בקובץ מכילה מזהה משתמש ברשת ואת המחיר (בדולרים) אותו הוא דורש.

## מטרה

בכל פעם שמשתמש רוכש בגד מחברת Praducci, הוא מכיר את החברה בהסתברות 1. כפי שמתואר לעיל, בתום ששת הסיבובים, ההסתברות שמשתמש יכיר את החברה אם הוא לא רכש בגד שווה ל-  $\frac{1}{1+10 \cdot e^{-\frac{B_t}{2}}}$ . המטרה שלכם היא למקסם את תוחלת כמות המשתמשים שמכירים את החברה. פורמלית, עליכם למקסם את:

$$\text{Objective function} = \mathbb{E}(\sum_{v \in V} I\{v \text{ knows Praducci after round 6} \mid \text{Influencers}\}) \quad (1)$$

כאשר  $V$  היא קבוצת הקודקודים (משתמשים) בגרף. התוחלת נלקחת על פני הסתברויות הרכישה השונות בהתאם למשפיענים שבחרתם, ועל פני הסתברויות החשיפה של כל משתמש לחברה. זוהי פונקציית המטרה שלכם, ואותה תרצו למקסם.

**שימו לב!** לנוחותכם, מצורף לקבצי התרגיל, קובץ קוד "Praducci\_simulation.py" המבצע סימולציה של המשימה כולה ומחזיר לכם את התוצאה שלכם. שימו לב כי התוצאה שהסימולציה מחזירה היא ריאליזציה יחידה של משתנה מקרי, ומטרתכם היא מקסום התוחלת של אותו משתנה מקרי.

## הגשה

עליכם להגיש שלושה קבצים:

1. קובץ csv יחיד המכיל את המשפיעים שבחרתם. על הקובץ להיות מיוצר על ידי הפונקציה `influencers_submission` שנמצאת בקובץ `Praducci_simulation.py` המקבלת את שני מספרי תעודות הזהות שלכם ואת רשימת המשפיענים שבחרתם.
2. קובץ הסבר בפורמט pdf, ששמו `ID1_ID2.pdf`, באורך של **לכל היותר עמוד**, הכולל את ההסבר שלכם לבחירת המשפיענים. הקובץ צריך לפרט את הגישה שלכם לפתרון הבעיה. המטרה של הקובץ הזה היא לנסח בבהירות את תהליך ההתמודדות שלכם עם המשימה.
3. בנוסף עליכם להגיש את הקוד שבעזרתו פתרתם את התרגיל - יש להגיש קובץ `py` יחיד שכותרתו `ID1_ID2.py`.

יש לכלול את שלושת הקבצים הללו בקובץ zip ששמו `ID1_ID2.zip`. אין להוסיף תיקויות נוספות בתוך ה-`zip`!

**הערה 1:** בכל הקבצים שקיבלתם השורה הראשונה מתארת את שמות השדות, השדות מופרדים על ידי פסיק.

**הערה 2:** ניתן לייבא את הספריות הבאות בלבד:

```
import numpy as np
import networkx as nx
import random
import pandas as pd
import csv
```

יבוא של ספריות אחרות יגרור פסילה של תרגיל הבית!

## הערכה וציון

הערכת הסטודנטים תתבצע לפי פונקציית המטרה במשוואה (1). כיוון שאין ביכולתנו לחשב את התוחלת במפורש נבצע אלפי סימולציות עד זמן  $t = 6$  עבור כל קבוצה, כדי למצוא רווח סמך קטן ככל האפשר. ציון הקבוצה יהיה מבוסס על אותו רווח סמך, כאשר הציונים יהיו על בסיס תחרות בין הגשות הסטודנטים.

**שאלה:** אם הציון שלנו מבוסס על משתנה מקרי, איך נוכל להבטיח הצלחה?  
**תשובה:** ראשית, הגשה שתשיג תוצאה של יותר מ-1100 (בריצת הבדיקה שנבצע אצלנו) תזכה את המגישים בציון 80 לפחות. במקרה והגשתכם לא עברה את הסף, לא יתקבלו ערעורים בטענה שבריצת הקוד שלכם כן הצלחתם לעבור אותו.

שנית, פונקציית הרווח במשוואה (1) אינה משתנה מקרי. זה **קבוע** התלוי בבחירה שלכם במשפיענים. מהות התרגיל היא לבחור את המשפיענים בצורה נכונה וטובה יותר מאחרים. לכן, יצירתיות, השקעה ויעילות יבטיחו לכם הצלחה.