

רקע

בתור סטודנטים מצוינים בעלי ידע רב בקורס מודלים במסחר אלקטרוני, חנות אינטרנטית גדולה (טרמינל-ווי) המוכרת את מרכולתה ברשת החברתית "איסטגלאם" פנתה אליכם וביקשה שתייעצו ותסייעו לה בשיווק מוצרי merchandise (חולצות, כובעים, גרביים וכו', ממותגים עם לוגו, תמונה או חתימה) של אמנים שונים.

לצורך כך, קיבלתם שני מסדי נתונים:

- חברויות ברשת החברתית "איסטגלאם", המאפשרת למשתמשים להעלות תמונות (ובפרט, תמונות בהן המשתמשים לובשים בגדים ממותגים). בכל פעם שמשתמש רוכש מוצר של טרמינל-ווי, חבריו ברשת יכולים לראות זאת לאחר מכן.
- נתונים מאפליקציית המוסיקה "ספוטיפלי", המתארים את מספר ההשמעות של כל אמן ע"י כל משתמש. נתונים אלו יכולים לשפוך אור על מידת האהדה של משתמש לכל אחד מן האמנים.

משימה

עבור כל אמן ברשימה שתקבלו (כפי שמתואר בהמשך), עליכם לבחור 5 משפיענים (influencers) שיקבלו מוצר (כלומר, בגד ממותג) של אמן זה בחינם בזמן $t = 0$. כאשר משתמש מקבלת/קונה מוצר של אמן מסוים, החברים ברשת רואים את מאורע הקנייה ומחליטים בהסתברות מסוימת האם לקנות את המוצר גם כן. המטרה שלכם היא שכמה שיותר משתמשים יקנו את המוצר לאורך 6 תקופות זמן, כלומר בנקודות הזמן $t = 1 \dots 6$ (כולל).

אמנים מקודמים

האמנים אותם אתם צריכים לקדם משתנים מקבוצה לקבוצה, ונקבעים על פי תעודות הזהות שלכם. לקבצי התרגיל מצורף קוד (choose_artists.py) שבעזרתו תוכלו לגלות מהם המזהים של האמנים המתאימים לכם. אנא הקפידו לקדם את האמנים הנכונים!

הערות טכניות

בחלק זה נתאר את תהליכי קבלת ההחלטות של המשתמשים ומבנה הרשת.

הסתברות רכישה

ההסתברות שמשתמש תרכוש מוצר של אמן מסוים תלויה הן בחבריה ברשת אינסטגלאם והן במספר הפעמים ששמעה המשתמש את שירי האמן כפי שמופיע בקובץ spotify.csv. כל שורה בקובץ זה מכילה את מספר הפעמים שכל משתמש שמע כל אמן. השדה הראשון – מזהה משתמש, השדה השני – מזהה האמן והשדה השלישי – מספר ההשמעות. אם לדוגמה לא קיימת רשומה המתארת את כמות הפעמים שאליס שמעה את האמנית עדן בת-צעיר, סימן שאליס לא שמעה מעולם את שירי האמנית.

את המוצרים 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 את המוצר 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 המוצר עד זמן t , **לא כולל**.
- נסמן ב- h את מספר הפעמים שהמשתמש האזין לשירי האמן.

ההסתברות לרכישה:

עבור משתמש שלא האזין מעולם לשירי האמן, ההסתברות הינה $\frac{B_t}{N_t}$.

עבור משתמש שהאזין לשירי האמן, ההסתברות הינה $\frac{h \cdot B_t}{1000 \cdot N_t}$.

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

מבנה הגרף והיווצרות קשתות

מבנה הרשת הינו דינמי, ותהליך היווצרות הקשתות אינו ידוע לנו בצורה מפורשת. כחוקרים חישוביים, תוכלו להיעזר במבנה הגרף בשתי נקודות הזמן הקודמות ($t = 0, -1$) כדי למדל את תהליך ההיווצרות.

אתם צריכים להסיק כיצד קשתות הולכות להיווצר!

הקובץ `instagram0.csv` מתאר את מצב הרשת בזמן $t = 0$. כל שורה בקובץ מכילה זוג מזהים של משתמשים. כלומר, שני משתמשים אלו הם חברים ברשת. שימו לב שחברות זוהי תכונה **הדדית** (כלומר גרף בעל קשתות לא מכוונות). הקובץ `instagram_1.csv` מתאר את מצב הרשת בזמן $t = -1$ על פי ייצוג דומה לזה של `instagram0.csv`. תוכלו להיעזר בשינוי שבין שתי נקודות הזמן כדי ללמוד את ההסתברות להיווצרות קשת.

דרך אחת לעשות זאת, למשל, היא להסתכל על קשתות שהוספתן לרשת בזמן $t = -1$ תסגור משולש (כפי שראינו בהרצאה ובתרגול, triadic closure), ולחשב את פרופורציית הקשתות שנוצרו בזמן $t = 0$.

הערות:

- הניחו כי קשתות בגרף לא יכולות להימחק, אלא רק להיווצר.
- ההסתברות להיווצרות קשת בין שני משתמשים תלויה, בין היתר, במספר חבריהם המשותפים ברשת החברתית.

הגרפים הבאים מתארים את החברויות בין אליס, בוב ואיב בזמנים t ו- $t + 1$:



כפי שניתן לראות בין זמן t לזמן $t + 1$ נוצרה חברות בין בוב לאיב.

נניח שבוב רכש את המוצר של האמנית נחל בוחבוט בזמן $t - 1$. אליס יכולה לראות זאת רק בזמן t ולא לפני. איב, לעומת זאת, יכולה לראות את הרכישה של בוב (מה ששיפיע על ההסתברות שהיא תרכוש את המוצר) רק בזמן $t + 1$.

הגשה

עליכם להגיש שלושה קבצים:

1. קובץ csv יחיד המכיל את המשפיעים שבחרתם עבור כל האמנים. על הקובץ להיות באותו פורמט בדיוק כמו זה של קובץ הדוגמא המצורף ושמו ID1_ID2.csv כאשר ID1 ו-ID2 הם מספרי תעודת הזהות של המגישים (עבור סטודנטים המגישים לבד שם הקובץ צריך להיות ID1.csv).
2. קובץ הסבר בפורמט pdf, ששמו הינו ID1_ID2.pdf, הכולל את ההסבר שלכם לבחירת המשפיעים. הקובץ צריך לפרט את הגישה שלכם לפתרון הבעיה. **הסבר לא מפורט עשוי לגרום לפסילת התרגיל.**
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 צריכים להימצא בקובץ zip ששמו ID1_ID2.zip. אין להוסיף תיקיות נוספות בתוך ה-zip!

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

הערה: ניתן לייבא את הספריות הבאות בלבד:

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

יבוא של ספריות אחרות יגרור פסילה של תרגיל הבית!

הערכה וציון

אנחנו נבצע מספר סימולציות אקראיות על המודל בעזרת המשפיענים שבחרתם עד לנקודת הזמן $t =$ 6. הציון שלכם יקבע ביחס למספר המשתמשים הממוצע שקנו את המוצר בסוף הריצה.

כאמור, האמנים אותם אתם צריכים לקדם יקבעו על פי מספרי תעודות הזהות שלכם. לקבצי התרגיל מצורף קוד (choose_artists.py) שבעזרתו תוכלו לגלות מהם המזהים של האמנים המתאימים לכם. אנא הקפידו לקדם את האמנים הנכונים!

הציון יהיה מבוסס על תחרות בין ההגשות של הסטודנטים.