

# אלגוריתמים ברשתות – דוח מסכם

## מגשים:

ליאור אקרמן 302346895

חי אליאסי 204726913

אורי פלה 314974858

נושא – רשת שיתופי פעולה בין אמנים ב-Spotify

מקור הנתונים - <https://www.kaggle.com/datasets/jfreyberg/spotify-artist-feature-collaboration-network/data>

לינק לפרויקט בDRIVE-

[https://drive.google.com/drive/folders/1-OijWb9ZboS4M9RjaP0PrMcP2jxToc\\_i?usp=sharing](https://drive.google.com/drive/folders/1-OijWb9ZboS4M9RjaP0PrMcP2jxToc_i?usp=sharing)

לינק לפרויקט בGITHUB-

<https://github.com/AcLior/Spotify-Artist-Feature-Collaboration-Network.git>

## מבוא –

מוזיקה היא חלק בלתי נפרד מהתרבות האנושית לאורך ההיסטוריה, והיא מהווה אמצעי לחיבור חברתי ולביטוי רגשי העידן הדיגיטלי הביא לשינויים מהותיים באופן שבו מוזיקה מופצת, צורכים אותה ומנתחים אותה. Spotify, אחת מפלטפורמות הזרמת המוזיקה הגדולות בעולם, מספקת נתונים ייחודיים המאפשרים ניתוח מבוסס נתונים של קשרי האמנים והמבנה החברתי של עולם המוזיקה.

בפרויקט זה נעסוק בניתוח רשת של אמנים ושיתופי פעולה בפלטפורמת Spotify, תוך שימוש בכלים של אלגוריתמים ברשתות לזיהוי מודלים של קהילות, מודולריות ומרכזיות. המטרה העיקרית שלנו היא לבחון כיצד ז'אנרים ומיקומים גיאוגרפיים משפיעים על מבנה הרשת ואופן התקשרות האמנים.

## **הכנת הנתונים**

בתחילת הפרויקט, התמקדנו ביבוא ובהבנה של הנתונים המגיעים מפלטפורמת Spotify. קיבלנו נתונים בפורמטים של קובצי CSV, ו-Excel כולל מידע על אמנים, שיתופי פעולה, ומאפיינים נוספים כמו ז'אנרים ומיקומים גיאוגרפיים.

## **ניקוי נתונים**

1. **הסרת ערכים חסרים:** נתקלנו בערכים חסרים במספר עמודות חשובות, כמו ז'אנרים ומיקומים גיאוגרפיים. במקרים מסוימים בחרנו להסיר את השורות עם ערכים חסרים, בעוד שבמקרים אחרים בחרנו למלא אותם על פי הגיון סטטיסטי או באמצעות ערכים נפוצים או ממוצעים.
  2. **המרת סוגי נתונים:** חלק מהעמודות היו דורשות המרה, למשל ממחרוזת של רשימות ז'אנרים לרשימה מעובדת של ז'אנרים בפייתון. זה דרש שימוש בפונקציות כמו `literal_eval` כדי להמיר את המחרוזות לאובייקטים.
  3. **סינון נתונים לפי תנאים מסוימים:** כדי להבטיח שהנתונים יהיו מדויקים יותר לניתוח, הסרנו נתונים שלא מתאימים לקריטריונים שהגדרנו, כגון סינון לפי מספר מינימלי של שיתופי פעולה או פופולריות מסוימת.
  4. **עיבוד עמודות נתונים מורכבות:** במקרים בהם עמודות הנתונים הכילו מידע מרובה שלא היה מעובד, כגון רשימות של מדינות או ז'אנרים, ביצענו פיצול והפרדה של הנתונים לפורמט נגיש יותר עבור ניתוחים.
- בחינת כיוונים לניתוח-** לאורך הפרויקט, ניסינו להשתמש בגישות שונות לניתוח הרשת, כולל שימוש באלגוריתמים שונים לזיהוי קהילות ומדדי מרכזיות. ניסינו אלגוריתמים כמו Girvan-Newman, ו-Louvain, וחקרנו את השפעתם על התוצאות.

**קשיים ומגבלות** - נתקלנו בכמה קשיים, כולל האתגרים שבזיהוי והסבר של הקשרים בין מדדי מרכזיות להצלחה במצעדים. בנוסף, חוונו קושי בניסוח שאלות מחקר ברורות מראש שיכוונו את הניתוח.

התמודדנו עם אתגרים שונים, כולל הקושי להסביר ממצאים ספציפיים או לקשר בין המודלים התיאורטיים לנתונים האמפיריים. דיונים אלו נעשו במטרה להבין את ההקשרים ולהעמיק את ההבנה של הממצאים.

ההתמקדות בכל שלב ובכל אתגר ספציפי שהועלה הובילה להעמקת ההבנה ולשיפור הגישות לניתוח רשתות גדולות ומורכבות.

### **שאלת המחקר**

**השפעת הז'אנר והמיקום הגיאוגרפי על מבנה הרשת ואופן התקשרות האמנים:**

- **מטרה:** להבין איך ז'אנרים ומיקומים גיאוגרפיים תורמים להיווצרות קהילות ברשת, ולזהות אם יש קבוצות שמבוססות על מאפיינים אלו.

**הקשר בין מידת המרכזיות של אמן להצלחתו במצעדים:**

- **מטרה:** לבדוק אם יש קשר בין מרכזיותו של אמן ברשת למשל (Degree Centrality) לבין מספר הלהיטים שלו במצעדים.

### **מהלך הפרויקט** –

הפרויקט נפתח בהגדרת המטרות והשאלות מחקר: לחקור את השפעת הז'אנר והמיקום הגיאוגרפי על מבנה הקהילות ברשת האמנים בספוטיפיי, ולבחון את הקשר בין מרכזיות האמן ברשת להצלחתו במצעדים. המחקר נערך על גבי דאטה של אמנים וקולבורציות בפלטפורמת ספוטיפיי.

טענו את הנתונים מקבצים שהכילו מידע על אמנים וקשרים ביניהם. נתקלנו באתגרים של נתונים חסרים ושגיאות בדאטה, שדרשו מאיתנו לפעול לניקוי מעמיק של הנתונים, כולל הסרת רשומות לא רלוונטיות ואיחוד קטגוריות שגויות.

לאחר ניקוי הנתונים, יצרנו גרף רשת בו הצמתים מייצגים אמנים והקשתות מייצגות קולבורציות. ביצענו חלוקה של הרשת לקהילות באמצעות אלגוריתם Louvain, כדי לזהות קבוצות של אמנים המתקשרים בתדירות גבוהה.

בחנו את הקהילות שנוצרו ואת מידת המרכזיות של כל אמן בקהילה שלו, וקשרנו זאת להצלחתם במצעדים. גילינו קשרים מעניינים בין מידת המרכזיות לבין מספר הלהיטים של האמנים במצעדים, וכן זיהינו קהילות שנוצרו על בסיס ז'אנרים ומיקומים גיאוגרפיים ספציפיים.

### **ניתוח קהילות-**

בהתבסס על החלוקה של הגרף לקהילות באמצעות אלגוריתם Louvain, זיהינו כי הקהילות נוצרות באופן טבעי לפי ז'אנרים ומיקומים גיאוגרפיים. למשל, קהילה אחת הורכבה בעיקר מאמני היפ הופ אמריקאיים, בעוד קהילה אחרת כללה אמנים מהז'אנרים האלקטרוניים מאירופה. ניתוח המודולריות נתן דירוג של 0.77 (לעומת דירוג של 0.7 Greedy Modularity), מה שמצביע על חלוקה אפקטיבית ומשמעותית של הרשת לקהילות חזקות ומובחנות.

### **קשר בין מידת מרכזיות להצלחה במצעדים**

הבדיקה המתמטית של הקשר בין מידת המרכזיות של האמנים ברשת לבין מספר ההופעות שלהם במצעדים העלתה קורלציה בינונית. לדוגמא, חישוב הקורלציה בין מידת הקרבה להצלחה הניב ערך של 0.35, מה שמרמז על קשר מובהק אך לא חזק במיוחד. הממצא הזה מעיד על חשיבות המיקום המרכזי של אמן ברשת להצלחתו המסחרית.

### **ניתוח תתי-קהילות**

בבדיקת תתי-קהילות בתוך קהילות מרכזיות גילינו שקיימים תת-קבוצות של אמנים המתקשרים באופן עז במיוחד. תוצאות זו מציינת את האפשרות לזהות מגמות תת-תרבותיות ואפילו שיתופי פעולה חדשים שעשויים להתפתח. באחת הקהילות, למשל, חברת אמני רוק גרמניים ופולק סקנדינביים, יצרו מרקם עשיר ומגוון של שיתופי פעולה.

### **ניתוח מודולריות לפי ז'אנר וגיאוגרפיה**

המודולריות לפי ז'אנר ולפי מיקום גיאוגרפי נבדקה כדי להבין עד כמה פרמטרים אלה תורמים למבנה הקהילות ברשת. החישובים הראו שיש גבולות ברורים ומוגדרים של קהילות לפי גיאוגרפיה וז'אנר, עם מודולריות גבוהה מעידה על חלוקה אופטימלית. לדוגמא, הקהילות שהיו מבוססות על ז'אנרים מובחנים

כמו פופ אמריקאי או קיי-פופ קוריאני הציגו מודולריות גבוהה מאוד, מה שמצביע על כך שז'אנרים אלו מהווים גורם מרכזי בקביעת אינטראקציות בתוך הרשת.

באופן דומה, נמצא שמיקום גיאוגרפי משפיע באופן ניכר על חלוקת הקהילות. למשל, אמנים מאירופה נמצאים בקהילות שונות מאמנים מצפון אמריקה, מה שמעיד על השפעה גיאוגרפית על דינמיקת הרשת. הניתוח הזה מאפשר לנו להבחין בתבניות של שיתוף פעולה והתקשרות תרבותית ומוזיקלית על רקע גיאוגרפי.

**דין-** הנתונים שנאספו וניתוחם הביאו להבנה עמוקה יותר של כיצד רשתות חברתיות בעולם המוזיקה מתפקדות ומשפיעות זו על זו. הממצאים מצביעים על כך שהתעשייה יכולה להיעזר במודלים אלו כדי לתכנן קמפיינים שיווקיים, ליצור שיתופי פעולה חדשים, ולהבין טוב יותר את הגורמים המשפיעים על הצלחה בתחום. ההמלצה היא להמשיך ולהעמיק במחקרים עתידיים שיבדקו גורמים נוספים המשפיעים על התנהלות רשתות בעולם המוזיקה.

ממצאים אלה מדגישים את החשיבות של שימוש בכלים אנליטיים מתקדמים וגישות חדשניות לניתוח נתוני רשת, ומציעים דרכים עבור חברות וארגונים בתעשייה לנצל את המידע הזה לקידום צמיחה והתפתחות. על ידי זיהוי הקהילות המשמעותיות והקשרים החזקים ברשת, ניתן להבין טוב יותר את המגנטים שמניעים את התרבות המוזיקלית ולפעול בהתאם לכך על מנת למקסם את הפוטנציאל של כל אמן או ז'אנר

## **סיכום** –

הפרויקט הביא להבנה עמוקה יותר של איך תכונות של ז'אנר וגיאוגרפיה משפיעות על מבנה הקהילות ברשת. גילינו כי קשרים חזקים בין אמנים בקהילות מסוימות נבנים על פי המיקום והז'אנרים המוזיקליים, מה שמדגים כיצד התרבות המוזיקלית מושפעת מגורמים אלו.

הקורלציות שמצאנו בין מדדי מרכזיות לבין הצלחה במצעדים סיפקו הצצה לאופן שבו המיקום המרכזי של אמן ברשת יכול להשפיע על קריירתו המוזיקלית. הפרויקט חשף גם על חשיבות החיבורים הבין-קהילתיים וההשפעה שלהם על יציבות ודינמיקה של הרשת כולה.

בסופו של דבר, המחקר עזר להבהיר כיצד ניתוח רשתות מוזיקליות מסוגל לספק תובנות חשובות לא רק על היחסים הקיימים אלא גם על הפוטנציאל לתכנן פעולות שיווק ופיתוח אומנים בעתיד.

בהתבסס על תוצאות אלו, מומלץ להמשיך ולפתח את כלי הניתוח וליישם אותם בפרויקטים נוספים שיעסקו בתחומים מוזיקליים ותרבותיים אחרים, לצורך העמקת ההבנה והשפעה על התעשייה המוזיקלית והתרבותית ברחבי העולם.

### **תרומה של כל אחד מחברי הצוות לפרויקט –**

#### **ליאור אקרמן:**

שימוש בכלים שנלמדו בקורס על הרשת כדי לבחון כיווני מחקר בחינת אלגוריתמי קהילה שונים והתאמתם למאפייני הרשת שלנו. כתיבת שאלות המחקר בניית המצגת והדוח המסכם

#### **אורי פלה וחי אליאסי:**

אחראים על איסוף הנתונים והכנתם, כולל ניקוי נתונים ראשוני. מחקר וכתיבה: כתיבת חלקים בדוח הסופי, כולל סקירת ספרות וניתוח מעמיק של הממצאים.