

# כריית נתונים ברשתות חברתיות ובמדיה הדיגיטלית בנושא נגיף הקורונה

מציגות: מלכה יחייס, עדי מבורך, שקד עוגן

מנחה אקדמית: ד"ר אירנה מילשטיין

## רקע

כיום, בעידן ה-BIG DATA, קיימים מאגרי מידע רבים. עלה הצורך לבצע ניתוח נתונים מחקרי וכריית מידע בנושא נגיף הקורונה מהרשתות החברתיות בדגש על "טוויטר" על מנת לגלות מידע חשוב על אופן העברת המסרים, תדירותם, בחינת קורלציות ואף גיבוש מסקנות אודותם.

## מטרות הפרויקט

1. בניית תמונת מצב לפי תוצאות ניתוח נתונים בנושא נגיף הקורונה מהרשת החברתית "טוויטר" על מנת לגלות מידע על אופן העברת המסרים, תדירותם, הקשרים ביניהם ומובילי דעה.
2. הסקת מסקנות בתחומים שונים (כלכלה, חברה, חינוך, בריאות וביטחון) בשיטת כריית טקסט.

## מתודולוגיה

שיטת המחקר הינה ניתוח ה-"Database" הנבחר ע"י עיבוד מקדים של הנתונים, סינון, סיווג ומיפוי ב-8 שלבים שונים, כל שלב כלל ניתוח סטטיסטיקה תיאורית. כתוצאה משלב הסיווג ומיפוי יצרנו 4 תתי מדגמים לחקירה נוספת ובחינת קורלציות ורגרסיה.

תת מדגם אוכלוסיית שציצה יותר מ-25 ציוצים בחודשיים

1 תת מדגם אוכלוסייה מרובת לייקים

תת מדגם אוכלוסיית הלא מאומתים (אנשים פרטיים)

תת מדגם אוכלוסיית המאומתים (אנשי ציבור, ידוענים)

בנוסף ניתוח טקסטואלי למילים מרכזיות נבחרות שעומדות בחזית המאבק בנגיף בקורונה, מה מידת השפעתם על האוכלוסייה והצגתם באופן ויזואלי בתוכנת NODEXL.

## ממצאים ומסקנות

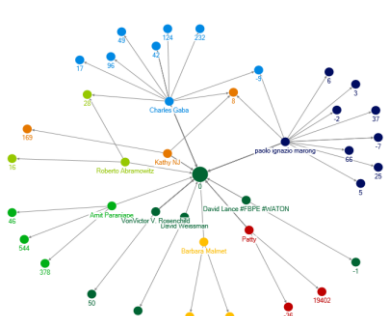
### אוכלוסיית המאומתים

- ככל שיש מספר עוקבים גבוה זה מחייב אותם לציץ יותר, על מנת לעניין את העוקבים ולשמור על כמות עוקבים גבוהה.
- משתמשים בעלי מספר חברים גבוה, מספר הציוצים קטן זאת לאור שאותם משתמשים מעדיפים לצרוך מידע מציוצים של חבריהם מאשר לציץ.
- מניתוח הטקסט זוהו בעיקר מילים חיוביות אודות נגיף הקורונה (1,206 מילים) לעומת (634 מילים) שליליות.
- מניתוח צמדי המילים העיקריות שנאמרו המילה Covid-19 ו-Cases חזרו על עצמם 107 פעמים ניתן להסיק מכך שבתקופה הנבחנת השיח הרוב עסק בעיקר במקרי תחלואה.

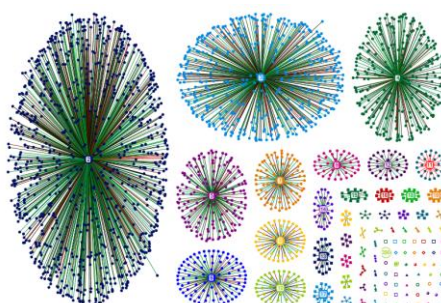
### אוכלוסיית לא מאומתים

- משתמשים מציצים יותר על מנת לגרוף עוקבים. משתמשים רבים "בטוויטר" מעוניינים להגדיל את מספר העוקבים שלהם ולהפוך למובילי דעת קהל.

### ניתוח תנודתיות כמות העוקבים



### כמות ציוצים ותכניה



### ענן מילים

