אלגוריתם: **clique percolation algorithm** - **אלגוריתם הפרקולציה של הקליקות**

**תקציר (Abstract)**

האלגוריתם מתאר גישה לזיהוי קהילות ברשתות גרפים על ידי חיפוש אחר k-קליקות, שהן תת-גרפים שלמים המכילים k צמתים. שתי k-קליקות נחשבות חופפות אם הן חולקות k-1 צמתים. קהילה מוגדרת כאיחוד מקסימלי של k-קליקות חופפות. זוהי גישה מקומית, המאפשרת חפיפה טבעית בין קהילות, ללא בעיית מגבלת רזולוציה.

היישום של הגישה מבוסס על איתור כל הקליקות המקסימליות במקום k-קליקות בודדות. זהו תהליך NP קשה, המשמעות היא שעבור רשתות גדולות, זמן הריצה יכול להיות אקספוננציאלי במספר הצמתים, למרות שרשתות עם מיליוני צמתים כבר נותחו בהצלחה בגישה זו.

**מבוא (Introduction)**

האלגוריתם Clique Percolation פותר בעיה בתחום חקר הרשתות, בעיקר בקונטקסט של רשתות חברתיות וביולוגיות. האלגוריתם מיועד לזיהוי קהילות או מודולים בתוך רשת גדולה יותר, כאשר קהילה מוגדרת כקבוצה של צמתים המחוברים היטב זה לזה, אך פחות מחוברים לצמתים מחוץ לקהילה. האלגוריתם מחפש קליקות (כלומר, קבוצות של צמתים בהן כל צומת מחובר לכל הצמתים האחרים בקבוצה) ומשתמש בהם כבניינים לבניית קהילות גדולות יותר.

היסטוריה קצרה של האלגוריתם וההתפתחויות העיקריות בתחום:

הרעיון של פרקולציה של קליקות נוצר במקור בהקשר של פיזיקה סטטיסטית ותורת הפרקולציה, אך התפתח לכלי חזק באנליזה של רשתות מורכבות. השימוש באלגוריתם לזיהוי קהילות ברשתות התחיל לקבל תשומת לב בתחילת המאה ה-21, עם עלייתה של תורת הרשתות המורכבות. הפרקולציה של הקליקות הוא אלגוריתם שהתפתח והשתכלל עם השנים תוך כדי שהוא משרת תחומים רבים ומתמודד עם דרישות מתמטיות וחישוביות מורכבות. הוא נחשב לכלי עוצמתי באיתור ובניית מודלים של קהילות ברשתות מורכבות, וממשיך להוות נושא למחקר ולהתפתחות.

* יישומים בתחומים נוספים: האלגוריתם החל להימצא לשימושים בתחומים נרחבים מעבר לחקר רשתות חברתיות, כולל ביולוגיה חישובית (לדוגמה, באנליזה של רשתות חלבונים), תקשורת, ואפילו בכלכלה.
* שיפורים אלגוריתמיים: בהמשך, פותחו שיפורים אלגוריתמיים שהקלו על זיהוי ואיחוד קליקות ברשתות גדולות, מה שהביא לשיפור ביעילות ובמהירות הזיהוי של קהילות.
* אינטגרציה עם טכניקות אחרות: האלגוריתם הוכלל במגוון רחב של כלים וספריות אנליזה של רשתות, ולעיתים משולב עם טכניקות ואלגוריתמים אחרים לצורך השגת תובנות מעמיקות יותר אודות מבנה ודינמיקה של רשתות.
* התמקדות במדדי רשתות: עם זמן, הפך הבנת האופן בו קהילות מושפעות ממדדים שונים של רשתות (כגון מרכזיות, דרגה, וקרבה) לחשובה יותר, והאלגוריתם הותאם לקחת בחשבון גורמים אלו בתהליך הזיהוי של קהילות.
* חקר השפעות חיצוניות: הוגדרה הצורך לחקור את השפעת גורמים חיצוניים כמו תקשורת, שינויים חברתיים, או אירועים ביולוגיים, על האופן שבו קהילות מתפתחות ומשתנות ברשתות, וכיצד אלגוריתם הפרקולציה של הקליקות יכול לזהות ולתאר את התהליכים הללו.

**הגדרות ומושגים יסודיים (Preliminaries and Basic Concepts)**

הגדרות ומושגים יסודיים באלגוריתם הפרקולציה של הקליקות:

רשתות חברתיות / ביולוגיות: מודלים של מערכות אינטראקטיביות המורכבות מצמתים (אנשים, חלבונים, וכו') וקשרים ביניהם.

קליקה (Clique): תת-גרף שלם, כלומר קבוצת צמתים שבה כל צומת מחובר לכל שאר הצמתים בקבוצה. בהקשר של האלגוריתם, מדובר ב-k-קליקה, שהיא קליקה של k צמתים.

פרקולציה (Percolation): תהליך שבו נוזל מסתנן או חודר דרך חומר מסונן. בהקשר של תורת הרשתות, הוא מתאר את התהליך שבו קהילות נוצרות דרך חפיפה רציפה של קליקות.

חפיפת קליקות: שתי k-קליקות נחשבות כחופפות אם הן חולקות k-1 צמתים. החפיפה מאפשרת את התהליך של פרקולציה ובניית קהילות.

קהילה (Community): בהקשר של האלגוריתם, קהילה מוגדרת כאיחוד מקסימלי של k-קליקות חופפות. זוהי קבוצת צמתים שהקשרים ביניהם חזקים יחסית לקשרים עם צמתים מחוץ לקהילה.

NP-קשה (NP-Hard): מונח במדעי המחשב המתאר בעיות שאין להן אלגוריתם יעיל שיכול לפתור אותן בזמן פולינומי. זיהוי קליקות מקסימליות הוא תהליך NP-קשה.

מגבלת רזולוציה: בעיה המתרחשת בכמה אלגוריתמים לזיהוי קהילות, שבה גודלו של הגרף מגביל את היכולת לזהות קהילות בגדלים מסוימים. אלגוריתם הפרקולציה של הקליקות מתגבר על בעיה זו על ידי התמקדות בחפיפת קליקות ולא בגודל הכולל של הקהילה.

**תיאור האלגוריתם (Algorithm Description)**

אלגוריתם הפרקולציה של הקליקות פועל במספר שלבים כדי לזהות קהילות ברשת גרף. התיאור הפורמלי והשלבים העיקריים של האלגוריתם מוצגים להלן:

הגדרת רשת: רשת G(V,E) כאשר V- הם הצמתים ו-E – הם הקשתות המחברים בין הצמתים.

הגדרת k-קליקה: k-קליקה מוגדרת כתת-גרף שלם של G המכיל K צמתים כך שכל זוג צמתים מחוברים זה לזה.

חיפוש קליקות: מציאת כל ה-k-קליקות ברשת.

חפיפת קליקות: שתי k-קליקות נחשבות כחופפות אם הן חולקות k−1 צמתים לפחות.

בניית קהילות: קהילה מוגדרת כאיחוד מקסימלי של k-קליקות חופפות.

**שלבי האלגוריתם:**

אתחול: קביעת ערך k, המספר המינימלי של צמתים בקליקה, כדי להתחיל את התהליך.

חיפוש קליקות: בדיקת כל תת-קבוצות האפשריות של צמתים בגודל k כדי לאתר k-קליקות. זהו תהליך שיכול להיות מאוד יקר במשאבים עבור רשתות גדולות.

הזיהוי של קליקות חופפות: לכל k-קליקה שנמצאה, בדוק את הקליקות האחרות שנמצאו כדי לזהות חפיפות של k−1 צמתים.

בניית קהילות: צמתים השייכים ל-k-קליקות חופפות מאוגדים יחד לקהילות. זהו השלב בו הקהילות השונות מוגדרות ונבנות מהקליקות החופפות.

A diagram of a network

Description automatically generatedאיחוד קהילות: כל קהילה שנמצאת היא איחוד מקסימלי של k-קליקות חופפות, וקהילות שונות יכולות לחפוף אם הן שותפות ל-k-קליקות.

**דוגמא להמחשה :**

* ראשית, נזהה רשת של צמתים וקשתות – (בתמונה רשת לא ממושקלת עם שמונה צמתים.)
* שנית, נזהה K קליקות שהן רשתות מחוברות במלואן עם k צמתים. ה k המינמלי האפשרי

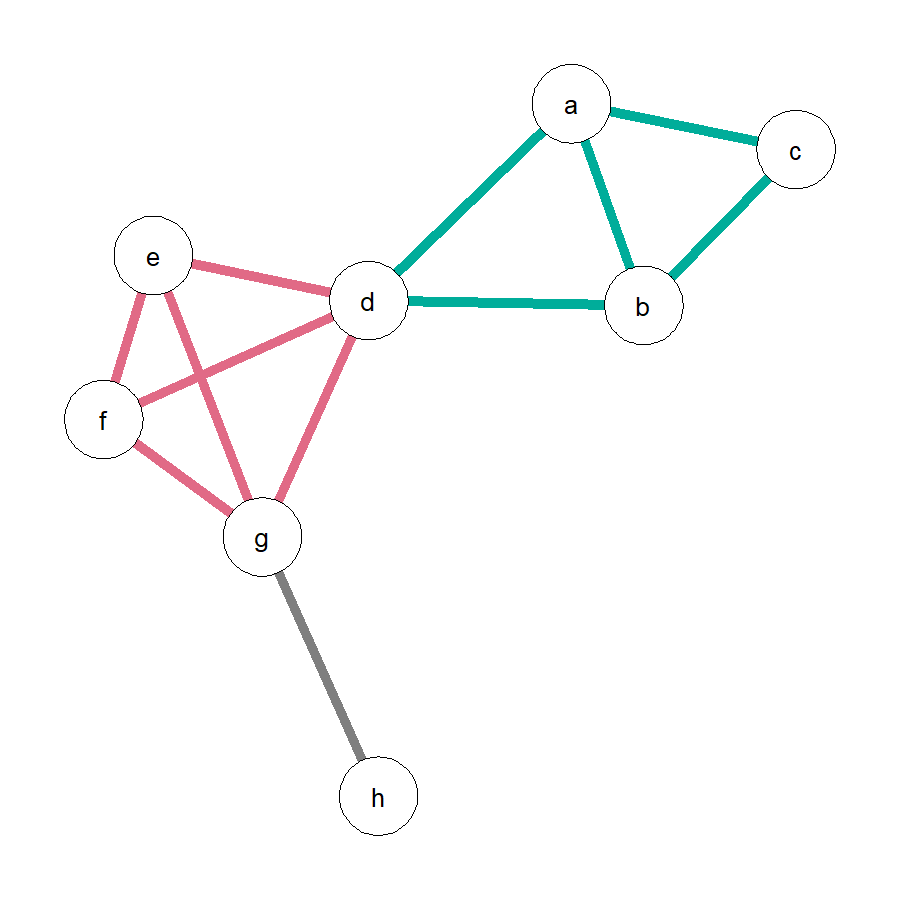
יהיה 3 אחרת הקליקות יהיו רק קשתות.

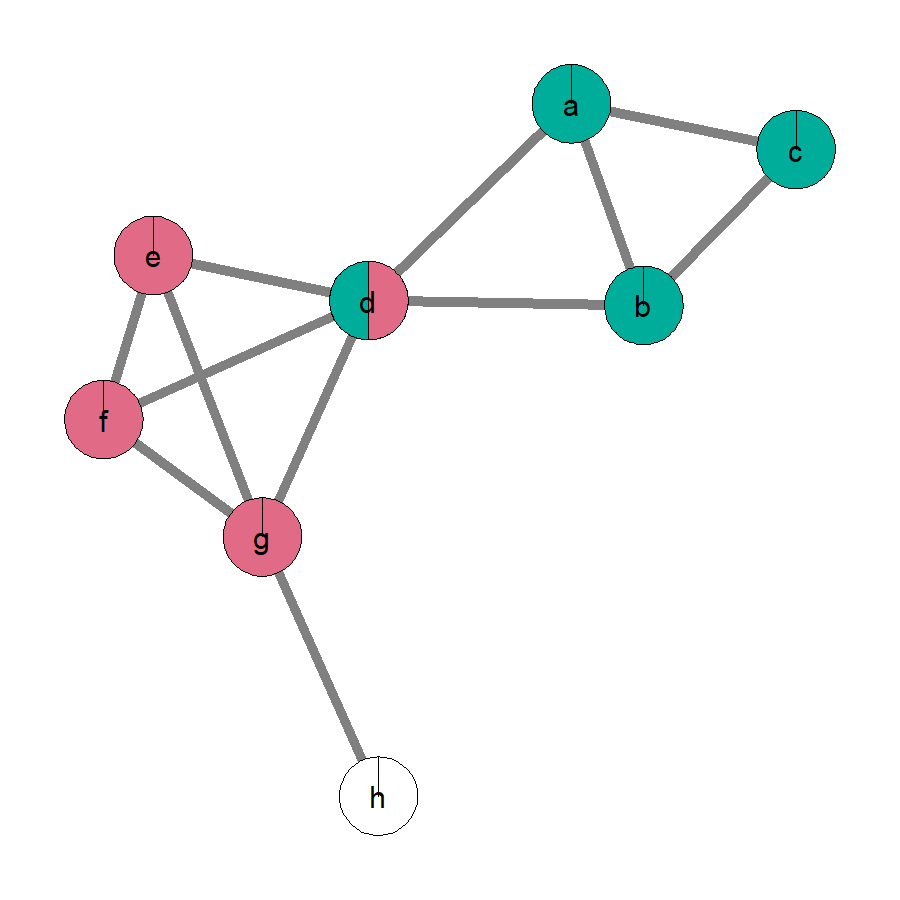
A group of lines and dots

Description automatically generatedהחלת k =3 על הדוגמה מובילה לזיהוי של שש k-cliques (k=3):

* לבסוף, קהילה מוגדרת כקבוצה של k קליקות צמודות,k קליקות שחולקות K-1 צמתים בדיוק. עם k=3, שתי 3-cliques צמודות אם הן חולקות בדיוק שני צמתים (שווה ערך לקשת). בדוגמה, זה מרמז שיש שתי קהילות (ראה להלן).

הקשתות הירוקים כוללים את3-cliques : a-b-c ו – a-b-d

הקשתות הוורודים כוללים את 3-cliques: d-e-f, d-e-g, d-f-g, e-f-g



* זה גם מראה שהאלגוריתם יכול להוביל לצמתים המשותפים לקהילות.

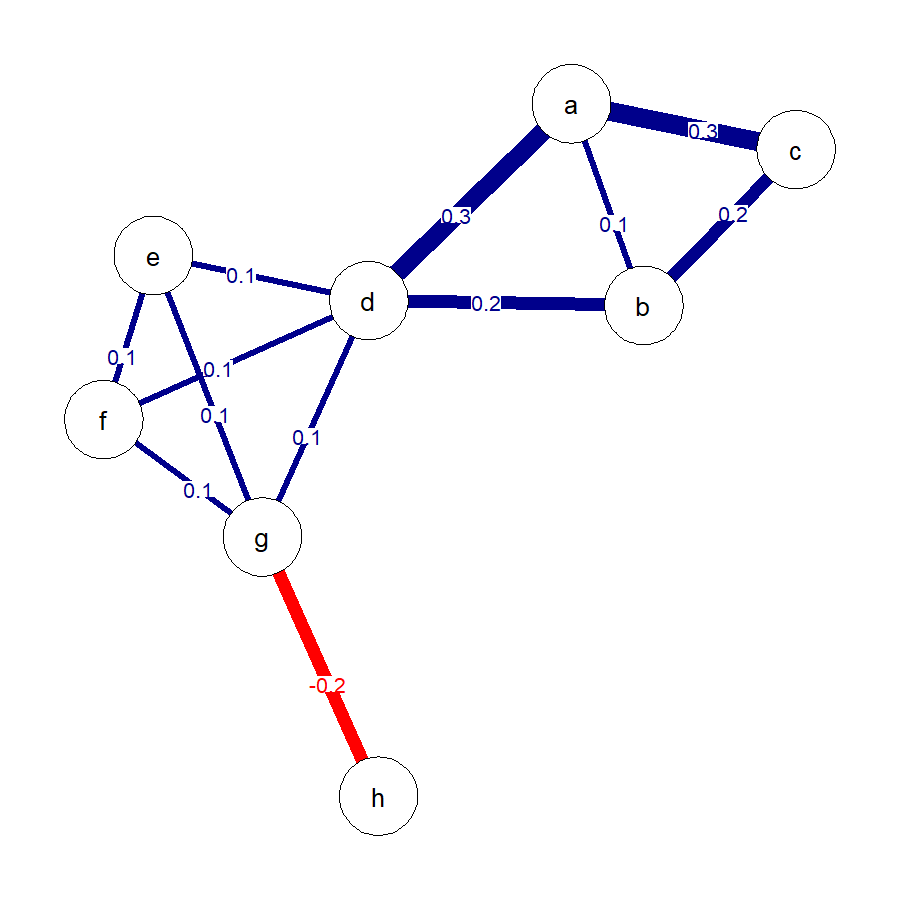
בדוגמה הנוכחית, צומת d הוא חלק מהקהילה הירוקה והורודה כאחד.

צמתים a, b ו-c הם חלק מהקהילה הירוקה בלבד והצמתים e, f ו-g הם חלק מהקהילה הוורודה בלבד.

חשוב לציין, האלגוריתם יכול גם להוביל לצמתים שאינם חלק מאף קהילה. אלה נקראים צמתים מבודדים, בדוגמה הנוכחית, צומת h.

עבור רשתות ממושקלות - לאלגוריתם יש רק שלב ביניים נוסף אחד. ספציפית, לאחר זיהוי ה k-cliques, הן נחשבות יותר להמשך רק אם העוצמה שלהן עולה על סף מוגדר I .

העוצמה של k-clique מוגדרת כממוצע הגיאומטרי של משקולות הקצה, כלומר:

**Ic=(∏(i<j;i,j∈C)wij)^2/k(k−1)**

כאשר:

C היא הקליקה,

i ו-j הם הצמתים,

w הוא משקל הקצה,

ו-k הוא גודל הקליק.

לדוגמה, 3 קליקות עם משקלי קצה 0.1, 0.2 ו-0.3 (בתמונה)

Ic=(0.1∗0.2∗0.3)2/3(3−1)≈0.18 (העוצמה של הקליקה היא 0.18)

A screenshot of a diagram

Description automatically generatedאם I=0.17, רק שתי קליקות (a–b-c, a–b-d) ישרדו את הסף. אם רק אלה ייחשבו עוד יותר, השיטה תמצא רק קהילה אחת (הקהילה הירוקה לשעבר) וארבעה צמתים יהיו מבודדים (e, f, g, h). עם זאת, אם I=0.09, כל הקליקים ישרדו את הסף, מה שיוביל לאותה חלוקה קהילתית כמו לרשת הלא ממושקלת.

**רכיבים עיקריים באלגוריתם:**

הגדרת k-קליקה וחיפוש קליקות: ליבת האלגוריתם, הדורשת חיפוש מעמיק ויסודי ברשת.

חפיפת קליקות ובניית קהילות: המפתח לזיהוי אוכלוסיות ברשת שמהווה קהילות טבעיות בתוך הרשת.

פתרון NP-קשה: אתגר חישובי שהאלגוריתם נאבק בו, מה שמצריך שימוש בטכניקות מתקדמות וייעול עבור רשתות גדולות.

**פסאודו קוד (Pseudocode)**

Algorithm Clique Percolation

Input: Graph G(V, E), Integer k

Output: List of Communities

1. Initialize an empty list of communities
2. Find all k-cliques in G and store them in list C
3. For each k-clique Ci in C:
   1. Initialize a new community with Ci
   2. For each k-clique Cj in C (where j != i):
      1. If Ci and Cj share (k-1) nodes:
         1. Merge Cj into the current community
         2. Remove Cj from C
4. Repeat Step 3 until no more cliques can be merged
5. Return the list of communities

**ניתוח סיבוכיות (Complexity Analysis)**

ניתוח סיבוכיות של אלגוריתם הפרקולציה של הקליקות דורש התייחסות לשני היבטים עיקריים: סיבוכיות זמן וסיבוכיות מרחב.

**סיבוכיות זמן:**

מציאת k-קליקות: השלב הראשון והכי זולל משאבים הוא מציאת כל ה-k-קליקות בגרף. מציאת קליקה בגודל k היא בעיה NP-קשה, ולכן הסיבוכיות שלה יכולה להיות אקספוננציאלית לגודל הגרף, במיוחד כאשר k גדול. במקרה הגרוע, ייתכן שנצטרך לבדוק כל תת-קבוצות של V בגודל k, מה שמוביל לסיבוכיות של ))

השוואת קליקות לזיהוי חפיפות: לאחר מציאת ה-k-קליקות, יש לבצע השוואות זוגות ביניהן כדי לזהות חפיפות. במקרה הגרוע, אם יש n k-קליקות, ייתכן שנצטרך לבצע השוואות.

איחוד קליקות לקהילות: תהליך האיחוד עשוי לכלול סריקות ואיחודים נוספים, אך סיבוכיות זו תלויה במספר הקהילות והחפיפות שנמצאו ובדרך כלל תהיה נמוכה יותר משלב החיפוש עצמו.

לכן, סיבוכיות הזמן הכוללת של האלגוריתם עשויה להיות אקספוננציאלית בהתאם לגודל הגרף ולערך k, במיוחד בשלב מציאת ה-k-קליקות.

**סיבוכיות מרחב:**

אחסון הקליקות: עלינו לאחסן את כל ה-k-קליקות שנמצאו, וכל קליקה יכולה להכיל עד k צמתים. אם n הוא מספר הקליקות שנמצאו, הסיבוכיות המרחבית לאחסון הקליקות היא O(nk).

אחסון הקהילות: בנוסף, יש לאחסן את הקהילות שנוצרו. המרחב הדרוש לכך תלוי במספר הקהילות ובגודלן, אך בדרך כלל לא יעלה על המרחב הדרוש לאחסון הקליקות.

בהתחשב בשני ההיבטים, סיבוכיות המרחב של האלגוריתם תלויה בעיקר במספר ובגודל הקליקות שנמצאו, אך היא נוטה להיות פולינומית בגודל הקלט.

סיבוכיות הזמן של האלגוריתם תיהיה אקספוננציאלית בהתאם לגודל הגרף ולערך k,  
בעוד שסיבוכיות המרחב נוטה להיות פולינומית, תלויה במספר ובגודל הקליקות שנמצאו.

**שימושים ויישומים (Applications and Use Cases)**

אלגוריתם הפרקולציה של הקליקות מצא שימושים ויישומים רבים בתחומים שונים כמו ביולוגיה, כלכלה, רשתות חברתיות ועוד, בשל היכולת שלו לזהות קהילות חופפות ומורכבות ברשתות גדולות. להלן כמה דוגמאות לשימושים של האלגוריתם בתחומים אלה:

**ביולוגיה:**

חקר רשתות חלבונים: האלגוריתם משמש לזיהוי קהילות של חלבונים המשתתפים בתהליכים ביולוגיים דומים או חלק ממולקולות מורכבות זהות. הבנת הקשרים בין חלבונים שונים והקבוצות שאליהן הם שייכים עשויה לסייע בפענוח תפקודים ביולוגיים ובפיתוח טיפולים חדשים.

אנליזה של רשתות גנטיות: באמצעות האלגוריתם ניתן לחקור אינטראקציות גנטיות ולזהות קבוצות של גנים המעורבים בתהליכים פיזיולוגיים דומים או במחלות מסוימות.

**כלכלה:**

ניתוח רשתות שוק ההון: האלגוריתם יכול להצביע על קהילות של מניות או נכסים פיננסיים המגיבים באופן דומה לשינויים בשוק, מה שיכול להציע הזדמנויות להשקעה או להבין טוב יותר את דינמיקת השוק.

בדיקת רשתות של חברות: חקר קשרים בין חברות עשוי לחשוף קהילות של חברות העובדות יחד או השותפות לאותם מגזרים או תחומים.

**רשתות חברתיות:**

ניתוח קהילות ברשתות חברתיות: האלגוריתם מאפשר לזהות קבוצות של משתמשים ברשתות חברתיות המשתפים עניינים דומים, תקשורת צפופה או קשרים חברתיים חזקים. זיהוי קהילות אלה יכול לשמש למטרות שיווק ממוקד, המלצות אישיות או פיתוח מוצר.

זיהוי תפקידים והשפעות של משתמשים מפתח: על ידי חקר האופן בו משתמשים מסוימים משמשים כגשרים בין קהילות, ניתן לזהות מנהיגי דעה או משתמשים בעלי השפעה מרכזית ברשת.

**ביקורת ואתגרים (Critique and Challenges)**

דיון ביתרונות וחסרונות של האלגוריתם.

הצגת אתגרים פתוחים ושאלות מחקר שנותרו.

**מחקרים מקריים (Case Studies)**

ניתוח של מספר מחקרי מקרה שבהם הופעל האלגוריתם והשפעתו על התוצאות.

**סיכום ומסקנות**