**דו"ח פרויקט-אלגוריתמים ברשתות**

**מגישים:**

בנימין חנונוב - 320472186

אושר שמיאן - 206389686

פיליפ שלייפר - 322357781

נושא-עבודת מחקר על רעידות אדמה בין השנים 1995-2023

מקור הנתונים - [Kaggle](https://www.kaggle.com/datasets/warcoder/earthquake-dataset)

לינק לפרויקט ב - [Drive](https://drive.google.com/drive/folders/12VxG3N8fqvcQlATrQNFcm-XeiMJXndkW?usp=sharing) (ניתן להריץ בענן בלי להוריד).

לינק לפרויקט ב - Github (לא ניתן להריץ בענן יש צורך להוריד) .

**מבוא –**

רעידות אדמה מהוות תופעה גיאולוגית בעלת השפעה ניכרת על חיי האדם והסביבה. בפרויקט זה, אנו עוסקים בניתוח רעידות אדמה שהתרחשו בין השנים 1995 ל-2023 תוך שימוש בטכניקות רשתות גיאוגרפיות. באמצעות ספריית NetworkX, אנו בונים גרפים דינמיים שבהם כל רעידת אדמה מיוצגת כצומת, והקשרים בין רעידות האדמה מחושבים לפי המרחקים ביניהן. באמצעות ניתוח זה, נחשוף את הקשרים בין רעידות האדמה, נבין את הדינמיקה שלהן, ונבחן שינויים במגמות לאורך השנים. הגרפים יאפשרו לנו לזהות דפוסים ולקבל תובנות חדשות לגבי התופעה, תוך בחינת השפעתם של מרחקים גיאוגרפיים ועל הקשרים ביניהן.

**הכנת הנתונים -**

בשלב הראשוני של הכנת הנתונים, זיהינו כי הדאטה כולל ערכים חסרים במספר תכונות, כמו 'continent' ו-country. כדי להבטיח את איכות הדאטה ולהקל על המשך הניתוח, ביצענו מספר פעולות עיבוד. ראשית, הסרנו את העמודות 'continent', 'country', ו-'location', אשר הוגדרו כלא רלוונטיות למודל הניתוח שלנו בעקבות הערכה מחודשת של השפעתן על מדדי המרכזיות ברשת.

לאחר מכן, ביצענו מיפוי מחדש לעמודות 'alert' ו-'magType', בהן הפכנו ערכים טקסטואליים לערכים נומריים, כדי לאפשר ניתוח סטטיסטי מתקדם ושילוב הנתונים במודלים מתמטיים בצורה מדויקת יותר. בנוסף, ערכנו עיגול לערכי הקואורדינטות הגאוגרפיות (Latitude ו-Longitude) במטרה לפשט את הנתונים, מה שסייע בהפחתת מורכבות החישובים ושיפור ניתוח הרשת. לבסוף, לאחר סיום העיבודים והטיפול בערכים החסרים, שמרנו את הדאטה הערוך לקובץ CSV חדש בשם "normalized\_data.csv" , המוכן לשימוש בשלבים הבאים של הניתוח. קוד העיבוד המלא, כולל הסברים מפורטים על כל שלב, מצורף בקובץ "data preparation.ipynb", אשר משמש כמדריך לתהליך הכנת הנתונים.

**קשיים ומגבלות -**

העבודה על ניתוח רעידות אדמה באמצעות גרפים דינמיים כללה מספר אתגרים משמעותיים. נתונים חסרים ולא אחידים דרשו התמודדות עם חוסרים וערכים לא עקביים, דבר שדרש זמן נוסף לעיבוד. הגדרת מטרות מדויקת הייתה קשה, והמעבר מגישה חיבורית לרעידות אדמה לגישה חדשה הצריך התאמה מהירה. יצירת גרפים דינמיים והבנת טכניקות גרפיות כמו חישוב מרחקים וייצוג קשתות היו מאתגרים. חישוב מדדים מרכזיים כמו קישוריות ומרכזיות דרש פיתוח קוד מותאם. ניסוי עם אלגוריתמים שונים לחלוקת קהילות ותיאום תוצאות בין אלגוריתמים שונים יצר אתגרים נוספים. הוויזואליזציה באמצעות Gephi ופרזנטציה ברורה של התוצאות דרשו הבנה מעמיקה של התוכנה וטכניקות ויזואליזציה. בנוסף, אינטגרציה בין חלקי הקוד השונים וההתמודדות עם נתונים לא אחידים הוסיפו מורכבות לעבודה.

**שאלת המחקר -**

במחקר זה נרצה לבדוק האם קיים קשר בין מדדי מרכזיות ברשת לנתוני רעידות האדמה. בנוסף, נרצה להבין האם ניתן להסיק מסקנות מהמציאות באמצעות ניתוח מדדי מרכזיות ברשת, והאם כדאי להשתמש במדדים ואלגוריתמים אלו לחקר הנושא לעומק. כמו כן, נחקור את הקהילות השונות ברשתות ונבחן מגמות לאורך מקבצי שנים.

**מהלך הפרויקט -**

שם הקובץ המכיל את קוד המחקר: "earthquake network analysis.ipynb" .

במהלך המחקר על רעידות אדמה, התחלנו את העבודה בגישה של ניסוי ותהיה כלומר, ניסינו לפתח את המתודולוגיה והטכניקות תוך כדי עבודה. בשלב הראשוני, התנסינו בגישה של חיבור רעידות אדמה באמצעות קווי מתאר. גישה זו נראתה לנו אינטואיטיבית, אך נתקלנו בקשיים בקביעת המטרה הנכונה והמדויקת. לאחר ביקורת בונה מהמרצה שלנו, חיפשנו גישה חדשה לעבוד על אותו הדאטה בצורה שונה .

בגישה החדשה, צומת מייצגת רעידת אדמה וכל קשת בין רעידות אדמה נושאת משקל שנקבע בהתאם למרחק בין הרעידות, כך שהמשקל הוא הפוך למרחק בקילומטרים כל עוד המרחק נמצא בטווח מסוים שנקבע מראש. כלומר, הקשתות מתמקדות בחיבור בין רעידות שנמצאות בטווח קרוב יחסית זו לזו, מה שאיפשר לחדד את הקשרים ולהציג בצורה ברורה יותר את הדינמיקה של רעידות האדמה. מכיוון שמה שהכי עניין אותנו ברשת הן רעידות האדמה שהן הצמתים בגרף רוב המדדים שנתמקד בהם יהיו ברמת הצומת.

במסגרת הניתוח שלנו, חילקנו את הנתונים לפי תקופות זמן מוגדרות כדי לנתח את התנהגות רעידות האדמה על פני השנים כך שהגדרנו ארבע תקופות עיקריות: 1995-2002, 2003-2010, 2011-2017, ו-2018-2023. לאחר מכן, פילחנו את הנתונים בהתאם לתקופות אלו וקיבלנו אותה כמות רעידות בערך לכל תקופת זמן. פילוח זה מאפשר לנו לעקוב אחרי השינויים בכמות וריכוז רעידות האדמה לאורך השנים ולבחון מגמות והשפעות בהתאם לתקופות השונות.

בתהליך יצירת הגרפים, התחלנו בהגדרת גרף חדש עבור כל תקופת זמן שנבדקה. לכל גרף הוספנו צמתים המייצגים את רעידות האדמה בתקופה זו. כל צומת מכילה מידע מקיף אודות הרעידה, כגון עוצמה, תאריך ושעה, פרמטרים גאולוגיים ומידע נוסף. בהמשך, הוספנו קשתות לא מכוונות בין הצמתים בהתבסס על המרחקים הגיאוגרפיים ביניהם. השתמשנו בנוסחת הוורסין (Haversine) לחישוב המרחק בין צמתי רעידה לרעידה , כאשר המרחק נמדד בקילומטרים. אם המרחק בין שני צמתים היה אפס (כלומר, רעידות זהות), הוגדרה קשת עם משקל של 1. אם המרחק היה פחות מ-1000 קילומטרים, הוגדר המשקל של הקשת כהפוך למרחק (1 חלקי המרחק). זאת כדי לשקף בצורה מדויקת את הקשרים בין רעידות אדמה קרובות ולהדגיש את הקשרים הקרובים יותר בין האירועים הגיאולוגיים. בסיום התהליך, כל גרף נשמר במילון המייצג את התקופות השונות, מה שמאפשר ניתוח והצגה ויזואלית של הקשרים בין רעידות האדמה לפי תקופות זמן מוגדרות.

לאחר שיצרנו את הגרפים עבור כל תקופת זמן, שמרנו את הגרפים בפורמט GraphML כדי לאפשר ויזואליזציה נוחה ויעילה באמצעות Gephi. לצורך כך, יצרנו תיקייה ייעודית לשמירת קבצי ה-GraphML, כך שנוכל להציג בדו"ח הפרויקט ויזואליזציה ברורה יותר .

הדבר הראשון שהחלטנו לבצע היה ניתוח של מדדי הדרגה של הצמתים בגרף. השתמשנו בפונקציה המובנית של הספרייה NetworkX כדי לחשב את דרגת הצמתים, שהיא סך הקשרים של כל צומת עם צמתים אחרים. חיפשנו את חמשת הצמתים עם הדרגות הגבוהות ביותר. לאחר מכן, בדקנו את דרגות הצמתים גם לפי אורכי המסלולים הקצרים ביותר ולא רק לפי מספר הקשתות, כדי להבין את הקשרים בין רעידות האדמה השונות. מה שמספק תובנות לגבי הפיזור והקשרים של רעידות האדמה.

מסקנות :

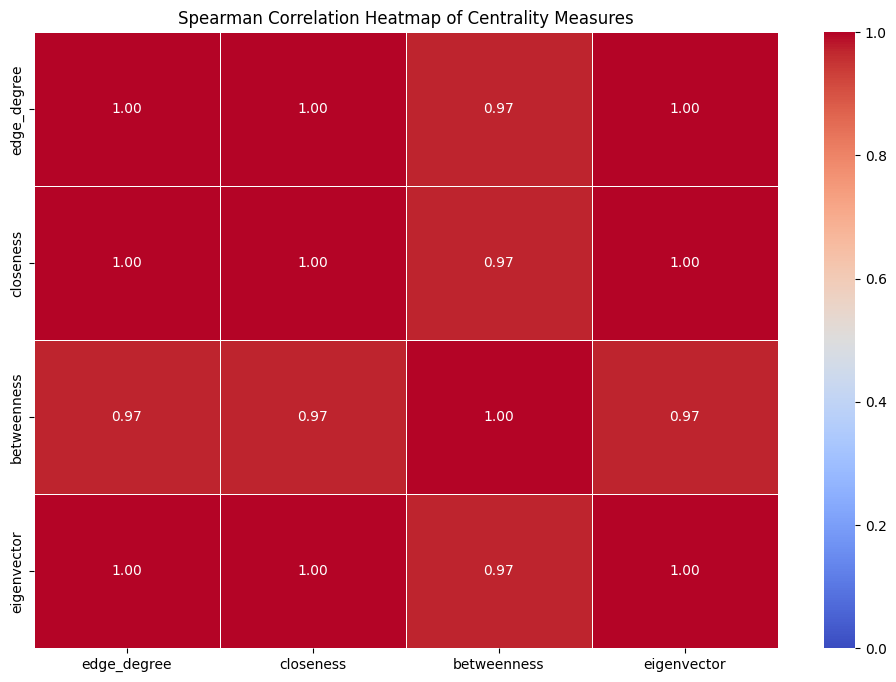
* **עלייה בממוצע:** הממוצע של דרגות הצמתים עלה בצורה משמעותית בין התקופות 1995-2002 ל-2003-2010, אך לאחר מכן נשאר כמעט קבוע עד 2017 וירד מעט בין 2018-2023.
* **התרחבות התפלגות:** דרגות הצמתים מציגות התפלגות רחבה יותר לאורך השנים, עם עלייה בערכים המקסימליים, שהגיעו ל-42 בתקופות האחרונות.
* **יציבות בשונות**: סטיית התקן יציבה יחסית ומעידה על יציבות .

בשלב הבא הוחלט לבצע ניתוח של מרכזיות ברשת על מנת להבין מי הצמתים הכי חשובים ברשת על פי המדדים הבאים :

* **Degree Centrality** : מדד המודד את מספר הקשרים הישירים של כל צומת ברשת.
* **Closeness Centrality**: מדד הבוחן את מרחקו של צומת מכל יתר הצמתים ברשת, ובכך מעריך את מידת הנגישות שלו.
* **Betweenness Centrality**: מדד הבודק את מספר הפעמים שצומת נמצא על הנתיב הקצר ביותר בין זוגות צמתים אחרים.
* **Eigenvector Centrality**: מדד המעריך לא רק את מספר הקשרים אלא גם את חשיבות הצמתים שאליהם צומת מקושרת.

על מנת לחלץ את המדדים האלו יצרנו קטע קוד שמחשב את מדדים שהוזכרו למעלה עבור כל רשת שיצרנו. לאחר מכן, בחרנו את הקודקודים בעלי הערכים הגבוהים ביותר על ידי פונקציה שמחזירה את n הקודקודים עם המדדים הטובים ביותר לכל מדד בצורה דינמית (ניתן להתאים את מספר הקודקודים המובילים לפי דרישה), תוך שמירה על ייחודיות בעזרת מבנה נתונים מסוג "set" המבטיח שכל קודקוד מוביל יופיע רק פעם אחת, ללא חפיפות בין המדדים השונים. זה מאפשר זיהוי מדויק של הצמתים החשובים ביותר ברשת בכל תקופת זמן מוגדרת.

לפני שנבדוק את הקורלציה בין מדדי המרכזיות לתכונות הדאטה המקורי, ביקשנו להבין את הקשרים בין מדדי המרכזיות עבור הצמתים הבולטים ביותר. הגרף מציג את הקורלציה בין מדדי המרכזיות השונים. ניתן לראות כי המדדים מקיימים קשרים חיוביים כמעט מושלמים זה עם זה, מה שמעיד על כך שצמתים בעלי Degree Centrality גבוהה נוטים גם להיות בעלי Closeness גבוה, Betweenness גבוה, ולפיכך, גם רמות Eigenvector גבוהות. הקורלציות בין המדדים מדגישה את ההשפעה הדומה של כל מדד על המרכזיות הכללית של הצמתים בגרף, כלומר, המסקנות לגבי מדד מרכזיות מסוים נוטה להיות תקפה גם לשאר המדדים.



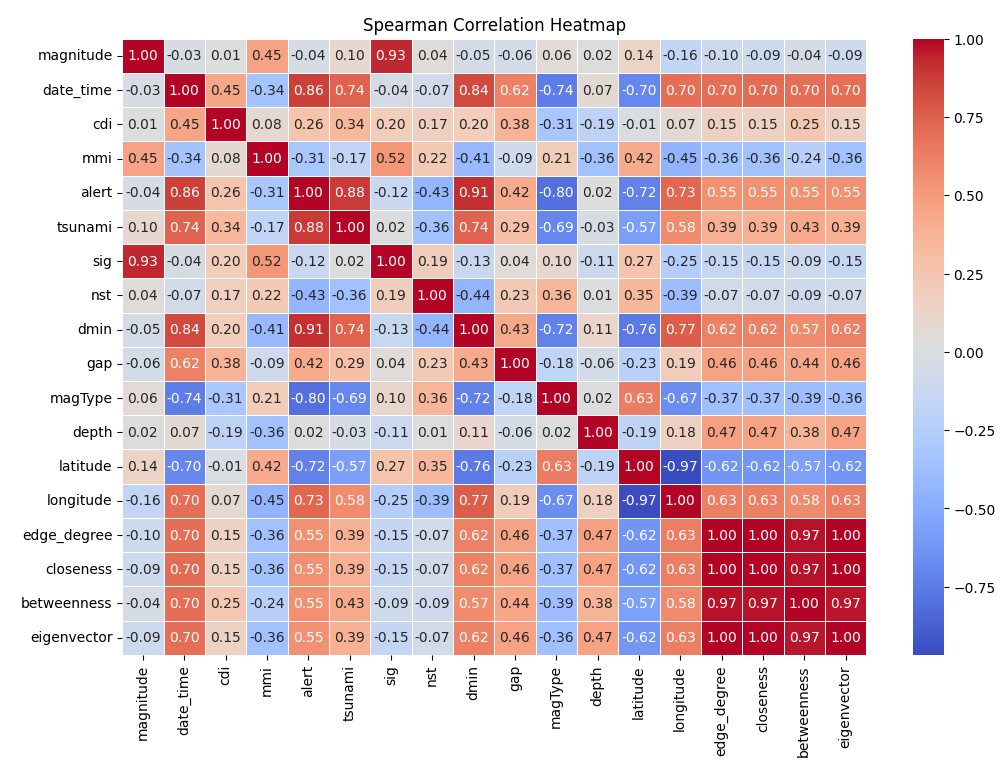
לאחר שזיהינו את הקודקודים המובילים על פי מדדי המרכזיות, רצינו לבחון אם קיים קשר בין תכונות רגילות מהדאטה שלנו, כמו עוצמת רעידת אדמה, מיקום גיאוגרפי ופרמטרים נוספים, לבין אותם קודקודים מובילים. לשם כך, בחרנו לבצע מבחן ספירמן על הקודקודים הללו בהתייחס למדדי המרכזיות בהם התמקדנו.

ביצענו מבחן ספירמן כדי לבדוק את הקורלציה בין מדדי המרכזיות לבין תכונות שונות בדאטה. הניתוח המתקבל מסייע לנו להבין את הקשרים הפוטנציאליים בין מדדי המרכזיות לבין התכונות השונות בדאטה, מטרת מבחן ספירמן כאן היא להבין אילו תכונות ספציפיות משפיעות על מדדי המרכזיות כך שיהיה ניתן להפיק תובנות נוספות על התנהגות הצמתים והקשרים ברשת.

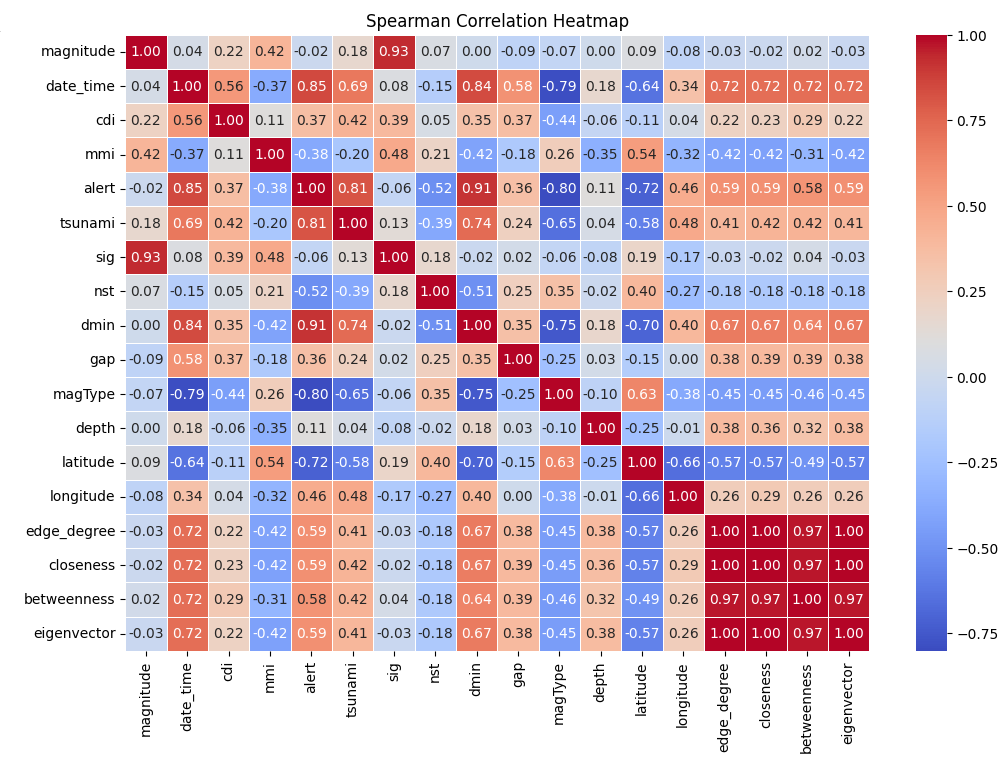
ביצענו הרצה לשני מקרים :

* n=5
* n=10

הרצה של n=5 :



הרצה של n=10:



**מסקנות ממבחן ספירמן מצביעות על מספר תובנות מעניינות:**

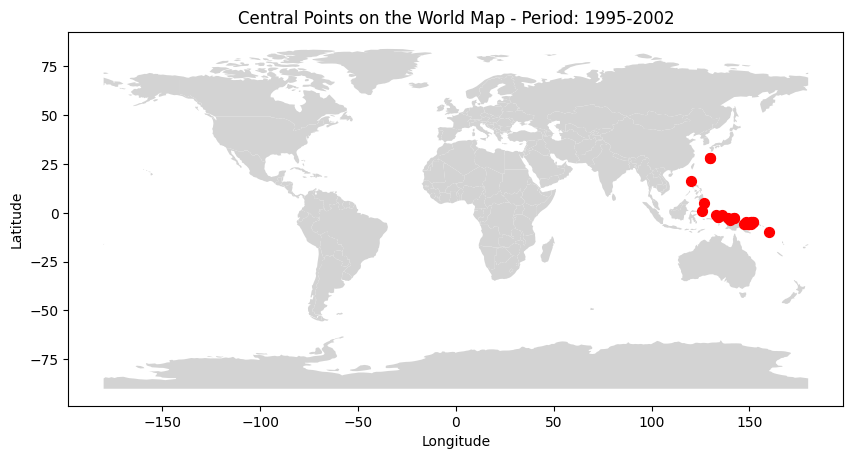
ניתן להבחין בכמה קורלציות חזקות, במיוחד בהתייחס למקום וזמן, מה שלא מפתיע לאור העובדה שאלה הן התכונות לפיהן יצרנו את הרשת. בולטת במיוחד השפעתו של המרחק הקרוב ביותר למרכז (dmin), אשר נראה שיש לו השפעה חשובה מאוד על מדדי המרכזיות. סביר להניח שהשפעה זו נובעת גם מצורתו הגיאוגרפית של האזור.

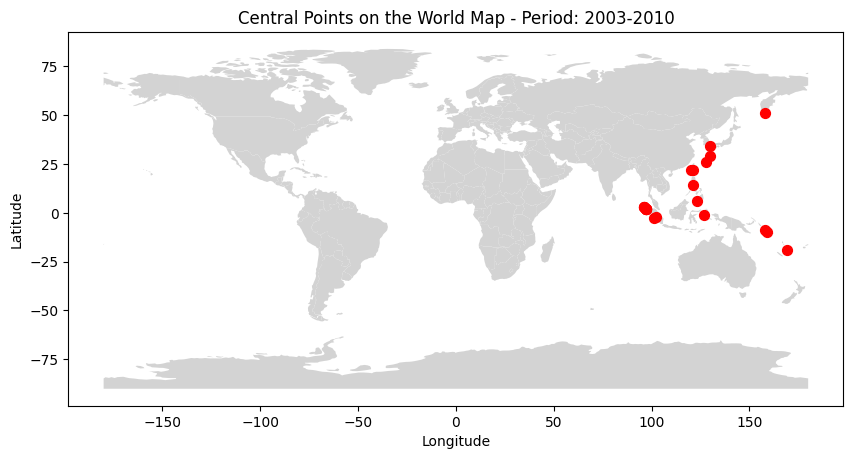
בנוסף, ישנה קורלציה בינונית עם עמודת האזהרה (alert), שיכולה להשפיע על תפקוד הצומת בתוך הגרף. מנגד, תכונות נוספות כמו mmi, sig, tsunami, ו-nst מציגות קורלציות חלשות או חלשות מאוד עם מדדי המרכזיות, ולכן הן אינן משפיעות באופן משמעותי על המרכזיות של צומת ברשת.

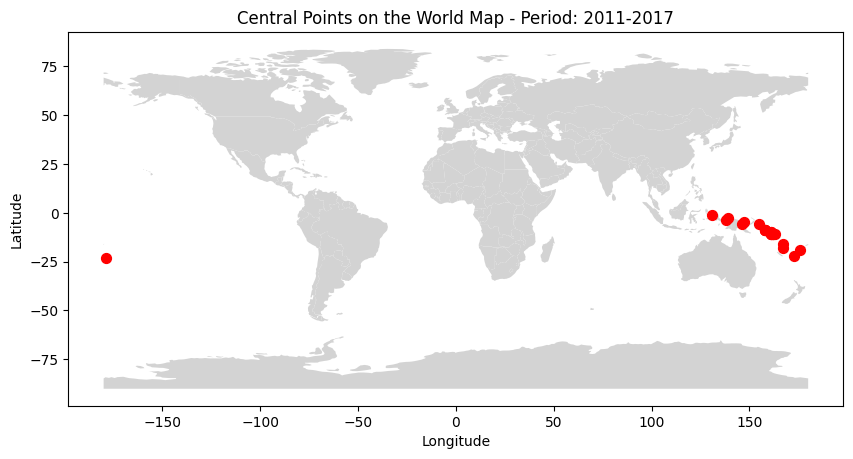
הנקודה המפתיעה ביותר היא שאין משמעות לעוצמת רעידת האדמה (magnitude) בהקשר של מדדי המרכזיות בגרפים שיצרנו, כלומר עוצמת הרעידה אינה משפיעה בצורה ישירה על המרכזיות של הצומת ברשת.

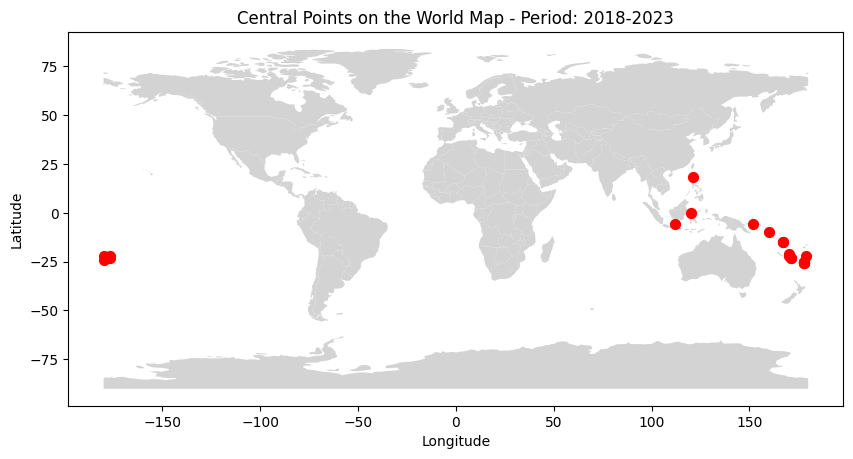
בהמשך, ניסינו להבין היכן ממוקמות רעידות האדמה (הצמתים) בעלות מדדי המרכזיות הגבוהים ביותר על פני מפת כדור הארץ. לשם כך, יצרנו קטע קוד שמציג את הצמתים המרכזיים הללו על המפה העולמית. באמצעות הוויזואליזציה הזו, ניתן לבחון האם ישנם שינויים במרכזיות הצמתים כאשר תקופות הזמן משתנות, וכך לקבל תובנות עמוקות יותר על הדינמיקה הגיאוגרפית של מדדי המרכזיות לאורך זמן.

**התפלגות הנקודות המרכזיות בעולם לפי תקופות**









**מסקנה:** באמצעות הוויזואליזציה שביצענו, ניתן לראות בצורה מובהקת שבמהלך כל התקופות שנבדקו, הצמתים בעלי מדדי המרכזיות הגבוהים ביותר ממוקמים באיי הפסיפיק הדרומיים. ממצא זה מצביע על כך שהאזור הזה מהווה מוקד מרכזי לרעידות אדמה לאורך זמן.

בהתבסס על המסקנה כי הצמתים בעלי מדדי המרכזיות הגבוהים ביותר ממוקמים באיי הפסיפיק הדרומיים, ניתן להרחיב את ההבנה ולבחון את הקשר בין מדדי המרכזיות לאזור רעידות האדמה המסוכן ביותר בעולם, הידוע בשם "טבעת האש" (Ring of Fire). אזור זה, המשתרע לאורך קצוות האוקיינוס השקט, מאופיין בפעילות סייסמית רבה, והוא מכיל את הריכוז הגבוה ביותר של הרי געש ורעידות אדמה. המרכזיות של הצמתים באזור זה מחזקת את ההשערה כי הוא מהווה מוקד סיכון עיקרי לרעידות אדמה לאורך זמן, ולכן חשיבותו הגיאולוגית היא משמעותית.



ביצענו ניתוח על פי קואורדינטות של רעידות אדמה כדי לקבוע אם כל רעידה נמצאת בתחום "Ring of Fire" (טבעת האש). לצורך כך, יצרנו פונקציה המדרגת רעידות אדמה לפי מיקומן הגאוגרפי ומוסיפה עמודה חדשה בוליאנית המייצגת האם כל רעידה נמצאת בתחום זה. לאחר מכן, חישבנו את אחוז הרעידות בכל תקופת זמן שנמצאות בתחום.

התוצאות היו מפתיעות, כאשר בכל אחת מהתקופות שנבדקו, כמעט כל הרעידות התרחשו בתוך גבולות "Ring of Fire". הפרופורציה היו חסרות תקדים.

להלן התוצאות המפורטות לכל תקופת זמן:

* 1995-2002: 86.86% מהרעידות נמצאו בתחום "Ring of Fire".
* 2003-2010: 92.68% מהרעידות נמצאו בתחום "Ring of Fire".
* 2011-2017: 90.60% מהרעידות נמצאו בתחום "Ring of Fire".
* 2018-2023: 91.82% מהרעידות נמצאו בתחום "Ring of Fire".

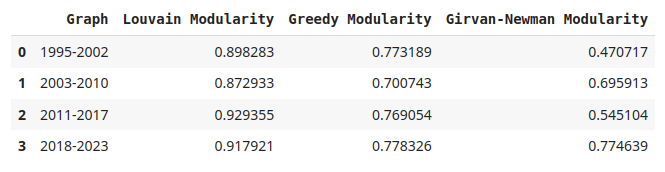
הנתונים מראים שברוב התקופות שנבדקו, אחוז גבוה של רעידות האדמה התרחש בתוך אזור "Ring of Fire". ממצא זה מחזק את ההבנה שיש קשר מובהק בין התרחשות רעידות האדמה לבין מיקומן הגאוגרפי של טבעת האש.

**קהילות -**

בעת ניתוח הקהילות רצינו להבין לאיזה אלגוריתם יש את רמת המודולריות הגבוה ביותר ולכן בחרנו להתמקד ב-3 סוגי אלגוריתם שלמדנו בקורס :

1. **אלגוריתם Louvain**: מבצע אופטימיזציה של מדד ה-modularity על ידי חילוק הרשת לקבוצות קהילות בצורה חכמה ושיפור התפלגות הקשרים בתוך הקהילות בהדרגה.
2. **אלגוריתם Greedy Modularity**: מחלק את הרשת לקהילות תוך כדי חיפוש מקסימלי של מדד ה-modularity על ידי מיזוג קהילות קיימות כדי להגדיל את הצפיפות של הקשרים בתוך הקהילות.
3. **אלגוריתם Girvan-Newman**: מסיר בהדרגה קשתות ברשת לפי מדד ה-edge betweenness כדי לחלק את הרשת לקהילות, ומבוסס על חישוב הקשרים החשובים ביותר לחיבור בין קהילות שונות.

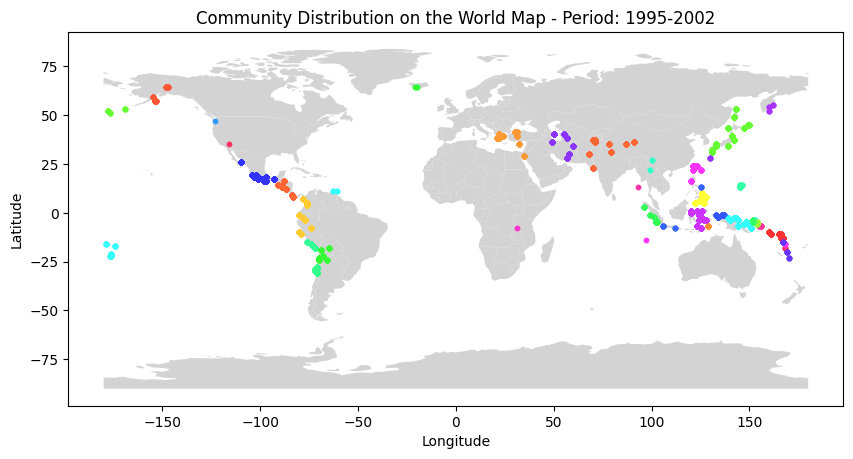
יצרנו קטע קוד שמבצע ריצה של שלושה אלגוריתמים שונים על כל גרף ומחשב את המודולריות של כל אלגוריתם. מטרת הקוד היא לזהות את האלגוריתם המתאים ביותר לחלוקת קהילות ברשתות גרפים המייצגות תקופות זמן של רעידות אדמה .

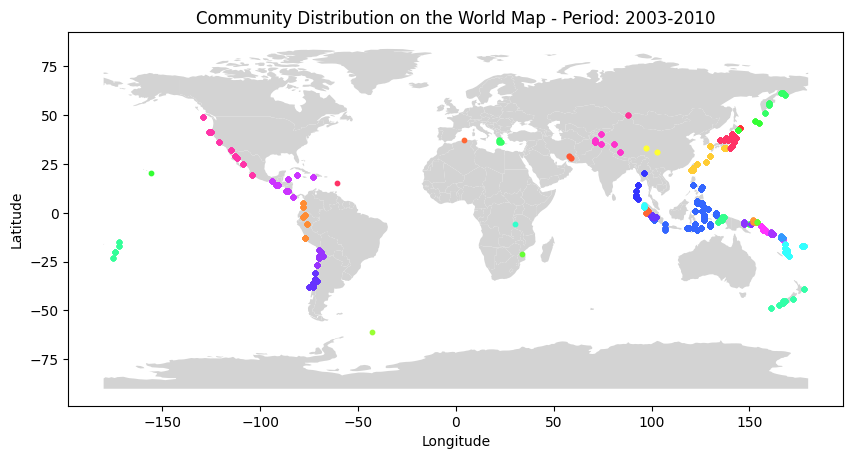


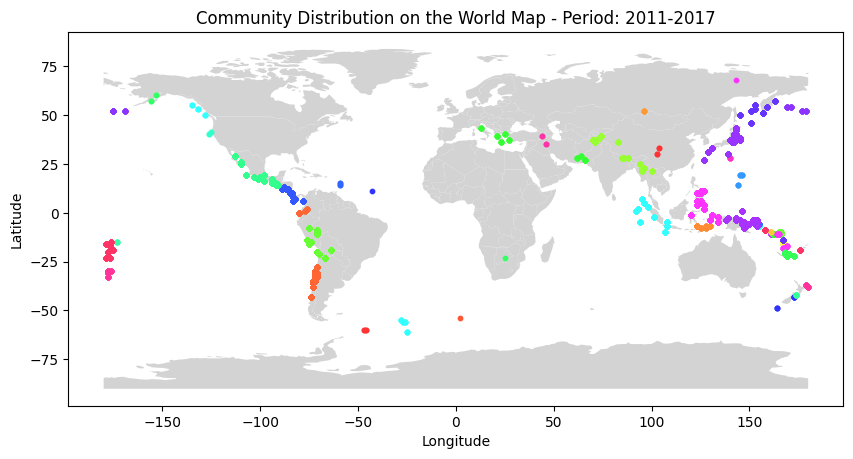
לאחר בדיקת ביצועי המודולריות של האלגוריתמים השונים לכל גרף, ניתן לראות שאלגוריתם **Louvain** הוא הטוב ביותר בצורה מובהקת בכל המקרים , על פי תוצאות המודולריות . בכך, חילקנו את הגרפים לקהילות בצורה אופטימלית, בהתאם לאלגוריתם שהפגין את הביצועים הטובים ביותר עבור כל תקופות הזמן.

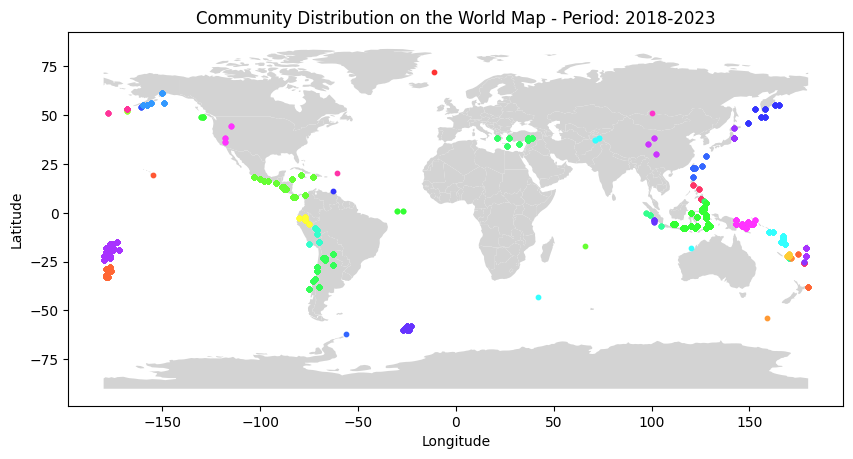
בהתבסס על חלוקת הקהילות הנוכחית שביצענו , ניתן לראות שהמצב הכללי של הקהילות נשאר דומה מאוד בין השנים השונות. בתקופת 1995-2002 ישנן 43 קהילות, בתקופת 2003-2010 יש 40 קהילות, ובתקופות 2011-2017 ו-2018-2023 יש 46 קהילות בכל אחת מהן .

במטרה להעמיק את הבנתנו במבנה הקהילות, יצרנו קטע קוד הממחיש את קהילות הגרף באמצעות ויזואליזציה על גבי מפה של כדור הארץ. בכל אחת מהקהילות, הצמתים מוצגים בצבע שונה, מה שמאפשר לזהות את הגבולות וההבדלים בין הקהילות בצורה אינטואיטיבית וברורה.









הממצאים מראים שאין שינוי דרמטי בגודל הקהילות בין התקופות השונות, מה שמעיד על יציבות יחסית במבנה הקהילתי לאורך הזמן. גם כאשר בוחנים את חלוקת הקהילות, המספרים נשמרים בקווים כלליים דומים, מה שמעיד על איזון ושימור של התפלגות הקהילות בתקופות הזמן השונות. זו תוצאה חיובית המצביעה על כך שהמפתחות החברתיים או המבניים שנמדדו לא השתנו באופן משמעותי לאורך התקופות שנבדקו.

בנוסף, הוויזואליזציה מדגישה בצורה מובהקת את תופעת ה-Ring of Fire בכל אחת מהתמונות שנבדקו. התופעה הזו, הניכרת כבר משנת 1995, ממשיכה להתקיים בעקביות עד לשנת 2023, סוף טווח הנתונים שלנו. הממצא הזה מצביע על רציפות והתמדה במבנה ובהתנהלות הקהילות לאורך שנים רבות, ומעיד על יציבות משמעותית במארג החברתי לאורך זמן.

השלב הבא היה יצירת קטע קוד שמסווג כל קודקוד בגרף לפי תקופת זמן ומוסיף לו את מספר הקהילה שאליה הוא משתייך. מטרתנו הייתה להבין אם קיימים קשרים בין השייכות לקהילה לבין מאפיינים אחרים בנתונים שלנו.

בהמשך לתצפיותינו על יציבות הקהילות, ביצענו ניתוח מעמיק כדי לבחון את הקשר בין מזהה הקלאסטר (cluster\_id) לבין משתנים נוספים בנתונים. הניתוח כלל שימוש במבחן ספירמן כדי לבחון את הקורלציה בין המזהה למשתנים אחרים. התוצאות לא הצביעו על קשר משמעותי בין מזהה הקלאסטר לבין יתר המשתנים הנבדקים. כלומר, אין השפעה ניכרת או קשר בולט בין ה-cluster\_id לבין פרמטרים חברתיים או מבניים אחרים, מה שמרמז על כך שמזהה הקלאסטר אינו משפיע ישירות על המאפיינים הנמדדים.

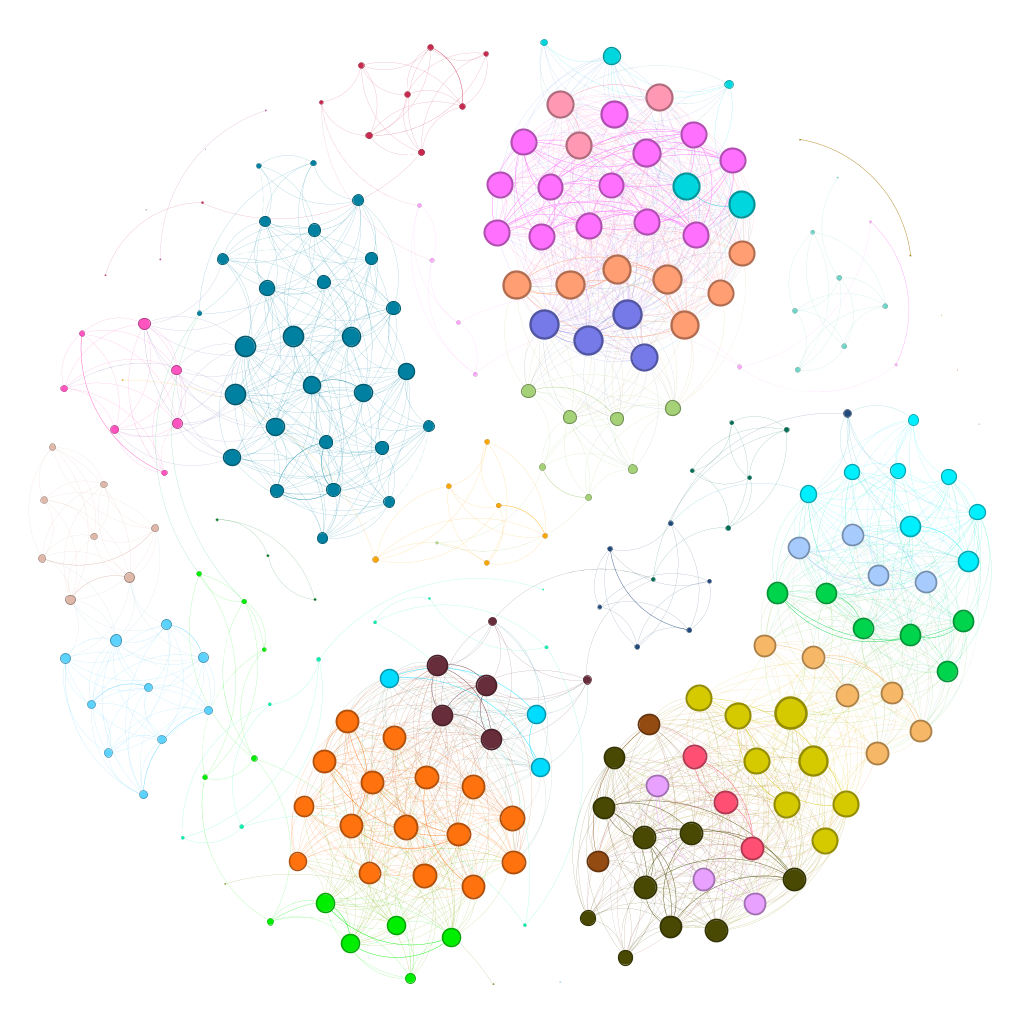
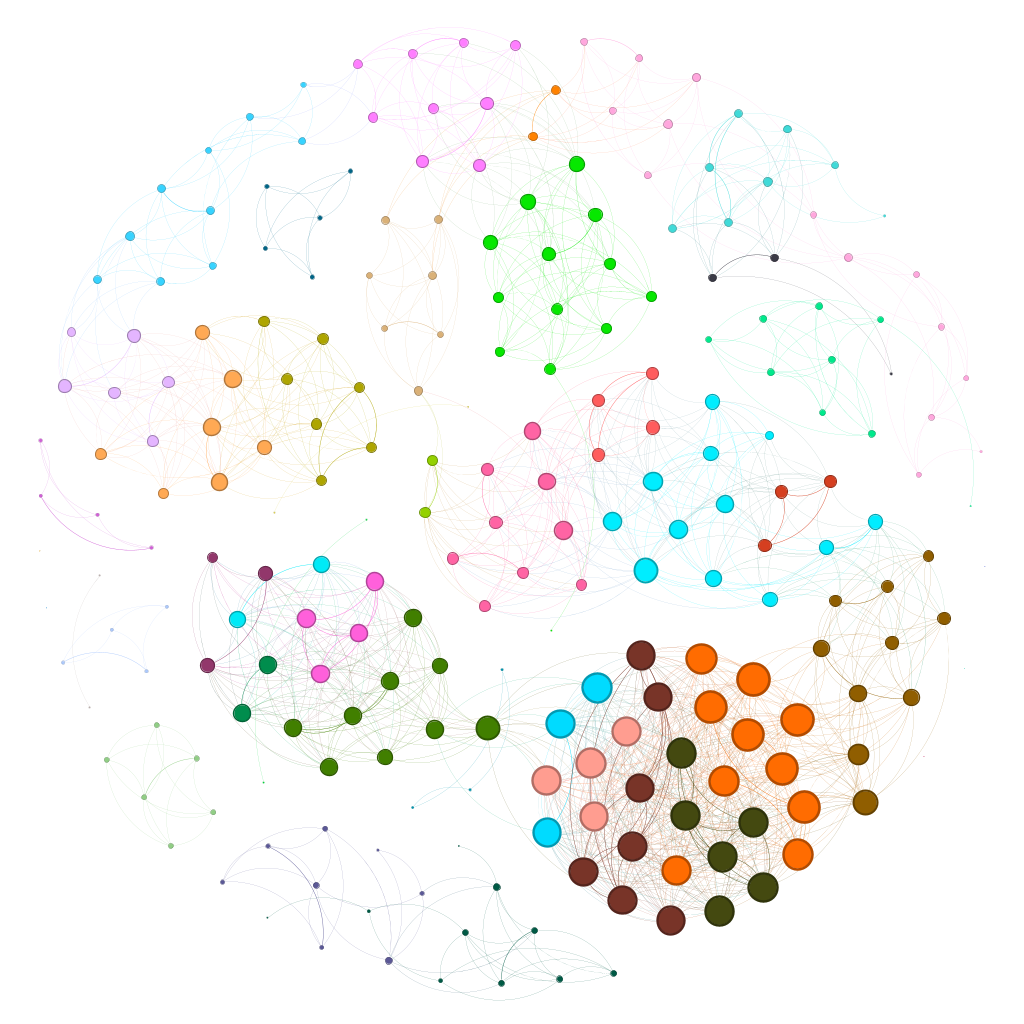
הממצאים מראים שיציבות הקהילות נשמרה לאורך השנים השונות. התקופות שנבדקו (1995-2002, 2003-2010, 2011-2017 ו-2018-2023) מצביעות על כך שהמבנה הכללי של הקהילות נותר עקבי, ללא שינויים דרמטיים. המספר הכולל של הקהילות נשמר בקווים כלליים דומים, מה שמעיד על יציבות ושימור ההתפלגות הקהילתית לאורך הזמן. תוצאה זו משקפת יציבות במבנה הקהילתי ומצביעה על כך שהמאפיינים החברתיים או המבניים שנבדקו לא השתנו באופן מהותי לאורך התקופות השונות.

המשמעות של יציבות הקהילות הנמדדות עבור רעידות האדמה היא שהמאפיינים הקשורים לרעידות האדמה נותרו עקביים לאורך זמן, גם כאשר הניתוח לא חשף קשר משמעותי בין מזהה הקלאסטר לבין משתנים אחרים. יציבות זו עשויה להצביע על כך שההתרחשויות הגיאולוגיות והשפעתן על הקהילות לא השתנו באופן משמעותי במהלך התקופות הנבדקות. זה רומז שהמאפיינים שמובילים לרעידות אדמה, כמו תנועות טקטוניות או שינויים גיאולוגיים, לא גרמו לשינויים דרמטיים באופני הפצתן והשפעתן על הקהילות לאורך הזמן.

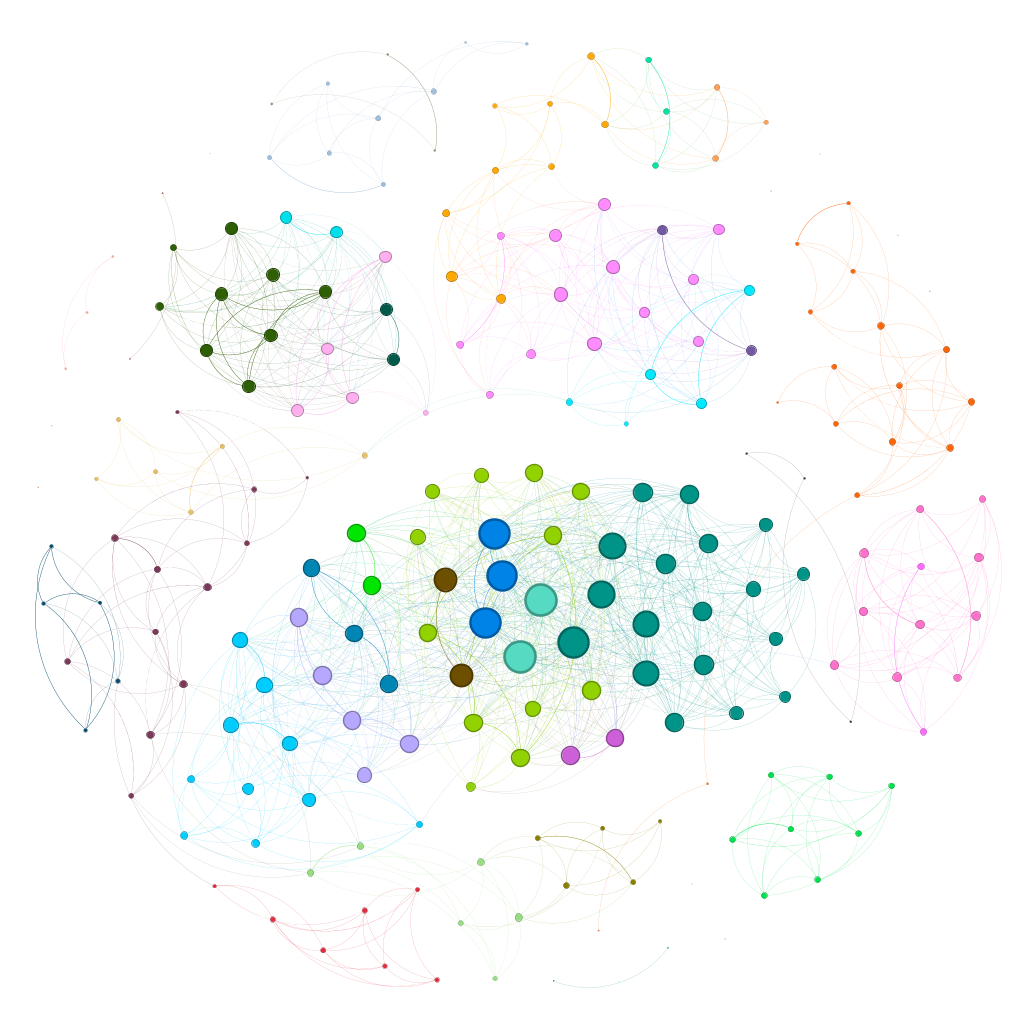
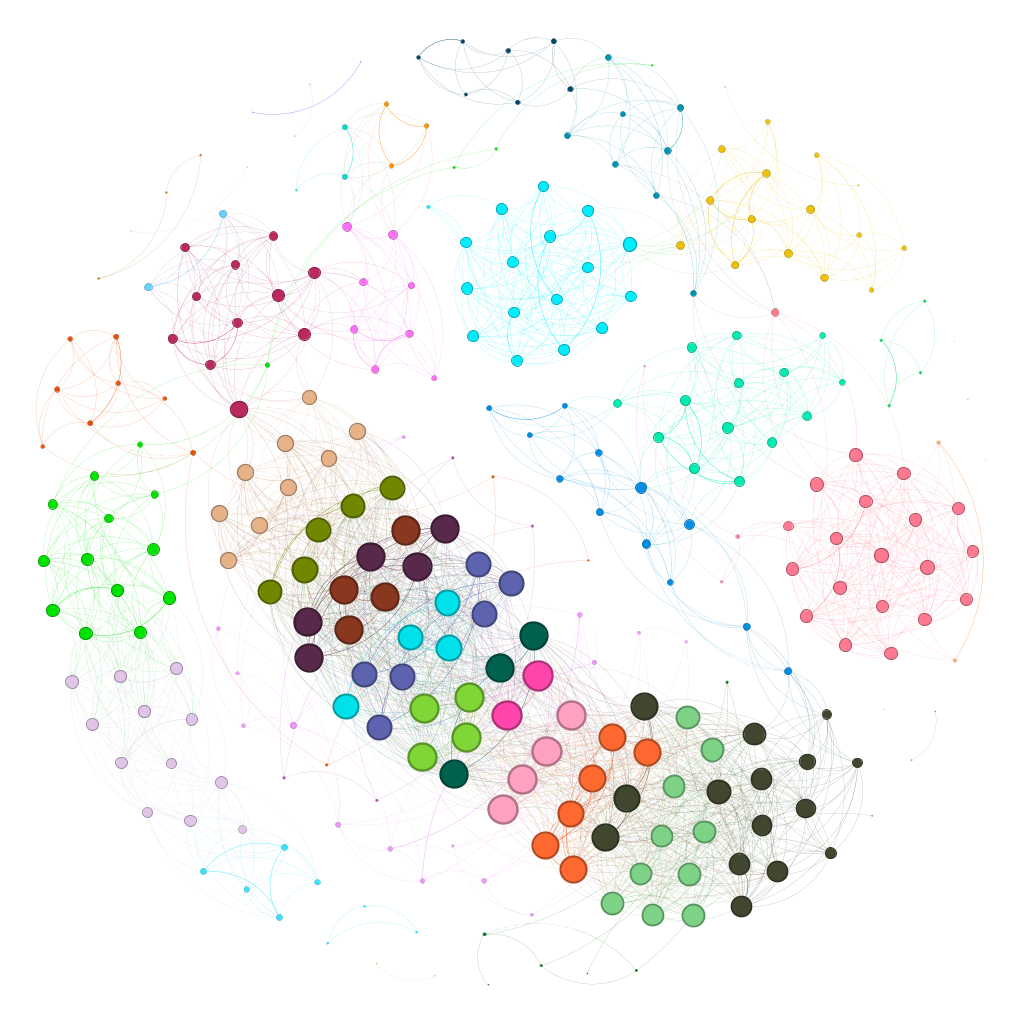
הוספנו ויזואליזציה המתארת את מיפוי הקהילות כדי להדגיש את הדמיון הרב בין הקהילות השונות. (התמונות בגודל המלא זמינות בתיקיית הדוח) .

**ויזואלית הקהילות :**

חלוקת הקהילות בוצעה באמצעות אלגוריתם Louvain, כפי שמתואר במחברת. לכל קהילה הוקצה צבע ייחודי עבור הצמתים, וגודל הצמתים הותאם לפי דרגת הצומת , כאשר הוא נע בין 1 ל-42. מדדי הקשתות שונו כך שצבע הקשת משתנה מבהיר לכהה יותר בהתאם לדרגת הקשת. קשתות עם דרגה גבוהה יותר יוצגו בצבע כהה יותר. תהליך זה בוצע באמצעות תוכנת Gephi .

1995-2002: 2003-2010:  


2011-2017 : 2018-2023:



**ניתוח הרשת הדינמית –**

בניתוח הרשת הדינמית של רעידות האדמה, אנו מתמקדים בהבנת הקשרים המשתנים בין רעידות האדמה לאורך תקופות זמן שונות בין השנים 1995 ל-2023. הרשתות שנבנות מייצגות את רעידות האדמה כצמתים, כאשר הקשרים בין הצמתים מתבססים על המרחקים הגיאוגרפיים ביניהם. במהלך הניתוח, כל תקופת זמן מוגדרת יוצרת גרף דינמי חדש, שבו קשתות בין הצמתים מקבלות משקל הפוך למרחק בין רעידות האדמה, כך שמרחקים קצרים יותר מקבלים משקל גבוה יותר. גישה זו מאפשרת לנו להבין כיצד רעידות האדמה הקרובות אחת לשנייה מקושרות זו את זו ברמות שונות, וכיצד הקשרים הללו משתנים במהלך השנים. באמצעות חישוב מדדי המרכזיות השונים, אנו יכולים לחשוף דפוסים עיקריים ולהבין את הדינמיקה של רעידות האדמה ברשת. ניתוח זה מספק תובנות על התנהגות הרשת במהלך הזמן, עוזר לזהות אזורים בעלי פעילות גיאולוגית גבוהה ומסייע לנו להבין כיצד רעידות האדמה משפיעות זו על זו במסגרת הרשת הגיאוגרפית הדינמית.

**דיון -**

בפרויקט זה התמקדנו בניתוח נתוני רעידות האדמה במטרה לזהות אזורים בעייתיים מבחינת סיכון לרעידות אדמה באמצעות אלגוריתמים של רשתות. השקענו את מירב המאמצים בניתוח המידע באמצעות מדדים של רשתות וקהילות, במטרה להבין את התפשטות התופעות והקשרים בין האזורים השונים. בהתבסס על הממצאים, אנו סבורים שיש מקום לשלב גם גישות נוספות, כמו ניתוח נתוני עתק (Big Data) ואנליזות מתקדמות נוספות, על מנת לשפר את הבנתנו בנוגע לאזורים הפגיעים ביותר. ייתכן שקיימות גישות נוספות ואלגוריתמים שונים בתחום רשתות שיכולים לתרום לתהליך זיהוי האזורים הבעייתיים בצורה מעמיקה ומדויקת יותר.

**סיכום -**   
  
בפרויקט זה, ניתחנו רעידות אדמה שהתרחשו בין השנים 1995 ל-2023 באמצעות טכניקות של רשתות גיאוגרפיות, תוך שימוש בספריית NetworkX. המטרה הייתה להבין את הקשרים בין רעידות האדמה באמצעות יצירת גרפים דינמיים, כאשר כל רעידת אדמה מיוצגת כצומת והקשרים בין הצמתים נקבעים לפי המרחקים הגיאוגרפיים ביניהם. השימוש ברשתות מאפשר לחשוף דפוסים ולהבין את הדינמיקה של רעידות האדמה, תוך ניתוח שינויים במגמות לאורך השנים ויכולת להבין את השפעתם של מרחקים גיאוגרפיים על הקשרים בין רעידות האדמה.

בשלב הכנת הנתונים, התמודדות עם ערכים חסרים ולא אחידים דרשה פעולות עיבוד כגון הסרת עמודות לא רלוונטיות, נרמול ערכים שונים, ועיגול של קואורדינטות גאוגרפיות. כל הנתונים עברו עיבוד לצורך ניתוח מעמיק. התהליך כלל גם טיפול באתגרים שונים כמו התמודדות עם קשיים בהגדרת מטרות, יצירת גרפים דינמיים, ויזואליזציה באמצעות Gephi, והבנה של טכניקות גרפיות מתקדמות.

במהלך ניתוח הרשת הדינמית, מדדנו את מדדי המרכזיות של הצמתים בגרפים שנוצרו עבור תקופות זמן שונות. ניתוח זה גילה עלייה בממוצע הדרגות של הצמתים במהלך התקופות הראשונות ולאחר מכן ירידה קלה. בדיקות נוספות גילו שהקשרים בין מדדי המרכזיות לבין עצמם היו כמעט מושלמים, מה שמצביע על יציבות יחסית בקשרים הפנימיים של הצמתים השונים.

בנוסף, ניתוח הקורלציה בין מדדי מרכזיות לבין תכונות רעידות האדמה הראה כי יש השפעה משמעותית למרחק מהמרכז הגיאוגרפי (dmin) ולקרבה לאזורי פעילות סייסמית על מדדי המרכזיות. לעומת זאת, לא נמצאה השפעה לעוצמת רעידת האדמה על המרכזיות של הצומת ברשת. ניתוח הקורלציה בנוסף חושף קורלציות בינוניות וחלשות בין משתנים שונים בין מדדי המרכזיות שמצביעים על תופעה שעלולה להיות חשובה במרכזיות הצומת בגרפים הדינימים. מיפוי התוצאות הגיאוגרפיות הדגיש את מרכזיות הצמתים באיי הפסיפיק הדרומיים, שהציגו דפוס מעניין המצביע על קשר מובהק לאזור ה-Ring of Fire. ניתוח זה חיזק את ההבנה של הקשר בין רעידות האדמה לבין אזור זה, הידוע כפעיל סייסמית.

לבסוף, ניתוח הקהילות באמצעות גישות שונות, כולל שיטות כמו Louvain, Greedy Modularity ו-Girvan-Newman, הוביל אותנו לבדוק את ביצועי המודליות של כל שיטה. ממצאי הניתוח הצביעו על כך שהשיטה של Louvain מציעה את חלוקת הקהילות האופטימלית ביותר עבורנו. התוצאות הראו שהמבנה הקהילתי נותר יציב יחסית לאורך השנים, ללא שינויים דרמטיים בגודל הקהילות או במיקומן. ממצאים אלו מעידים על יציבות בגיאוגרפיה של רעידות האדמה ובמארג החברתי שלהן לאורך זמן.

**תרומה של כל אחד מחברי הצוות לפרוקיט -**

**בני**: פיתח את הקוד לניתוח הרשת ושימוש ב-gephi.

**פיליפ**: ביצע את הניתוח הסטטיסטי והניתוח הדינמי של הרשת.

**אושר**: כתב את סקירת הספרות והציג את הממצאים וההמלצות.