

## נספחים

### נספח א' - אלגוריתם ליידן לזיהוי קהילות ברשת

אלגוריתם ליידן הוא אלגוריתם שנועד לזהות קהילות בתוך רשתות. האלגוריתם מחפש את החלוקה הממקסמת את המודולריות של הרשת. מודולריות מוגדרת באופן הבא:

$$M \equiv ActualShare - \gamma \cdot ExpectedShare$$

ה- $ActualShare$  מייצג את שיעור הנסיעות ברשת המתבצעות בתוך הקהילות ומחושב באופן פורמלי באמצעות הנוסחה הבאה:

$$ActualShare = \frac{\sum_c f_c}{F}$$

כאשר  $f_c$  הוא סך הנסיעות המתבצעות בתוך קהילה  $c$  ו- $F$  מייצג את כלל הנסיעות ברשת. ה- $ExpectedShare$  הוא תוחלת השיעור של הנסיעות בתוך הקהילות בחלוקה השומרת על מבנה הקהילות ברשת המקורית אך מחלקת את הנסיעות באופן אקראי. באופן פורמלי:

$$ExpectedShare = \sum_c \left( \frac{F_c}{2F} \right)^2$$

כאשר  $f_c$  מייצג את סכום הדרגות של הצמתים בקהילה  $c$ .<sup>1</sup> פרמטר הרזולוציה,  $\gamma$ , נקבע על ידי החוקר. כך, מדד המודולריות תחום מלמעלה ב-1. כאשר פרמטר הרזולוציה שווה 1 והמדד מקבל ערך קרוב ל-0, החלוקה הנבחנת טובה כמו חלוקה אקראית של הרשת לקהילות. ערך נמוך מ-0 מצביע על חלוקה טובה פחות מחלוקה אקראית של הרשת.<sup>2</sup> על מנת להמחיש זאת, נדגים את האופן בו האלגוריתם פועל על רשת פשוטה עם שלושה צמתים:

| סה"כ | א  | ב  | ג  | סה"כ |
|------|----|----|----|------|
| א    |    | 1  | 2  | 3    |
| ב    | 3  |    | 4  | 7    |
| ג    | 5  | 6  |    | 11   |
| סה"כ | 8  | 7  | 6  | 21   |
| דרגה | 11 | 14 | 17 | 42   |

השורות מתייחסות לאזורי מוצא ואילו העמודות לאזורי יעד. האלגוריתם אינו מתחשב בנסיעות פנימיות בתוך האזורים ועל כן האלכסון של הטבלה ריק. כעת נניח כי הרשת מחולקת לשתי קהילות – קהילה אחת הכוללת את צמתים א' ו-ב' וקהילה נוספת הכוללת את צומת ג' בלבד. הנסיעות המתבצעות בתוך הקהילות ברשת הן הנסיעות מ-א' ל-ב' (1) ומ-ב' ל-א' (3). ערך ה-*ActualShare* עבור חלוקה זו מחושב באופן הבא:

$$ActualShare = \frac{1 + 3}{21} = 0.19$$

כעת לשם חישוב ה-*ExpectedShare* נחשב תחילה את הדרגות של כל אחת מהצמתים. דרגה של צומת היא סכום הנסיעות היוצאות והנכנסות ממנה ואליה. לפיכך הדרגה של צומת א' היא 11, הדרגה של צומת ב' היא 14 ואילו הדרגה של צומת ג' היא 17. ה-*ExpectedShare* עבור החלוקה המוצעת הוא:

$$ExpectedShare = \frac{(11 + 14)^2 + 17^2}{(2 \cdot 21)^2} = 0.518$$

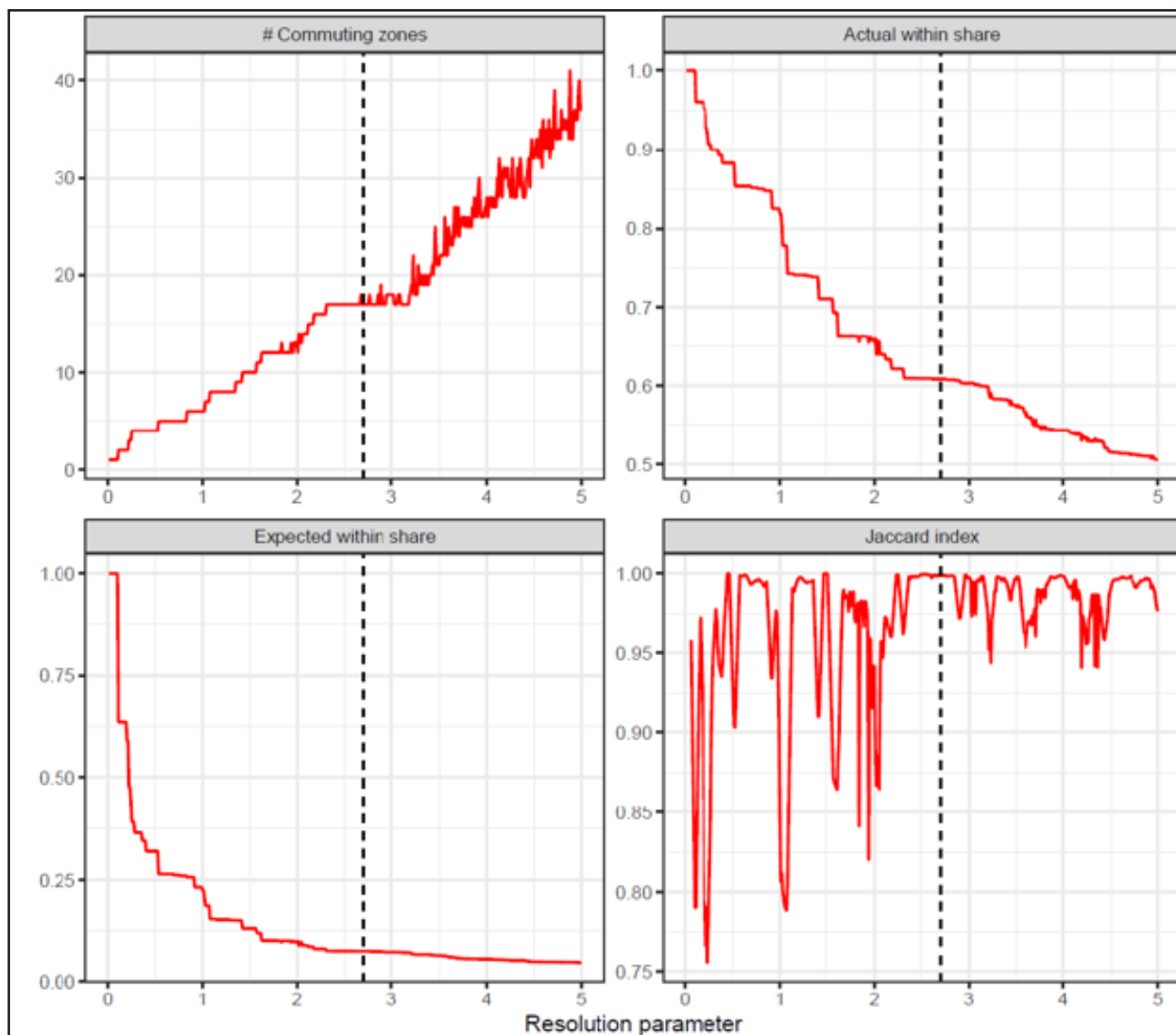
ערך המודולריות עבור קהילה זו תלוי בפרמטר  $\gamma$ . עבור  $\gamma=1$  נקבל:

$$M = 0.19 - 1 \cdot 0.518 = -0.328$$

משמע, החלוקה המוצעת לקהילות טובה פחות מחלוקה אקראית של הרשת. אלגוריתם ליידן נועד למצוא את החלוקה של הרשת בה ערך המודולריות הוא הגבוה ביותר עבור פרמטר הרזולוציה שנקבע.

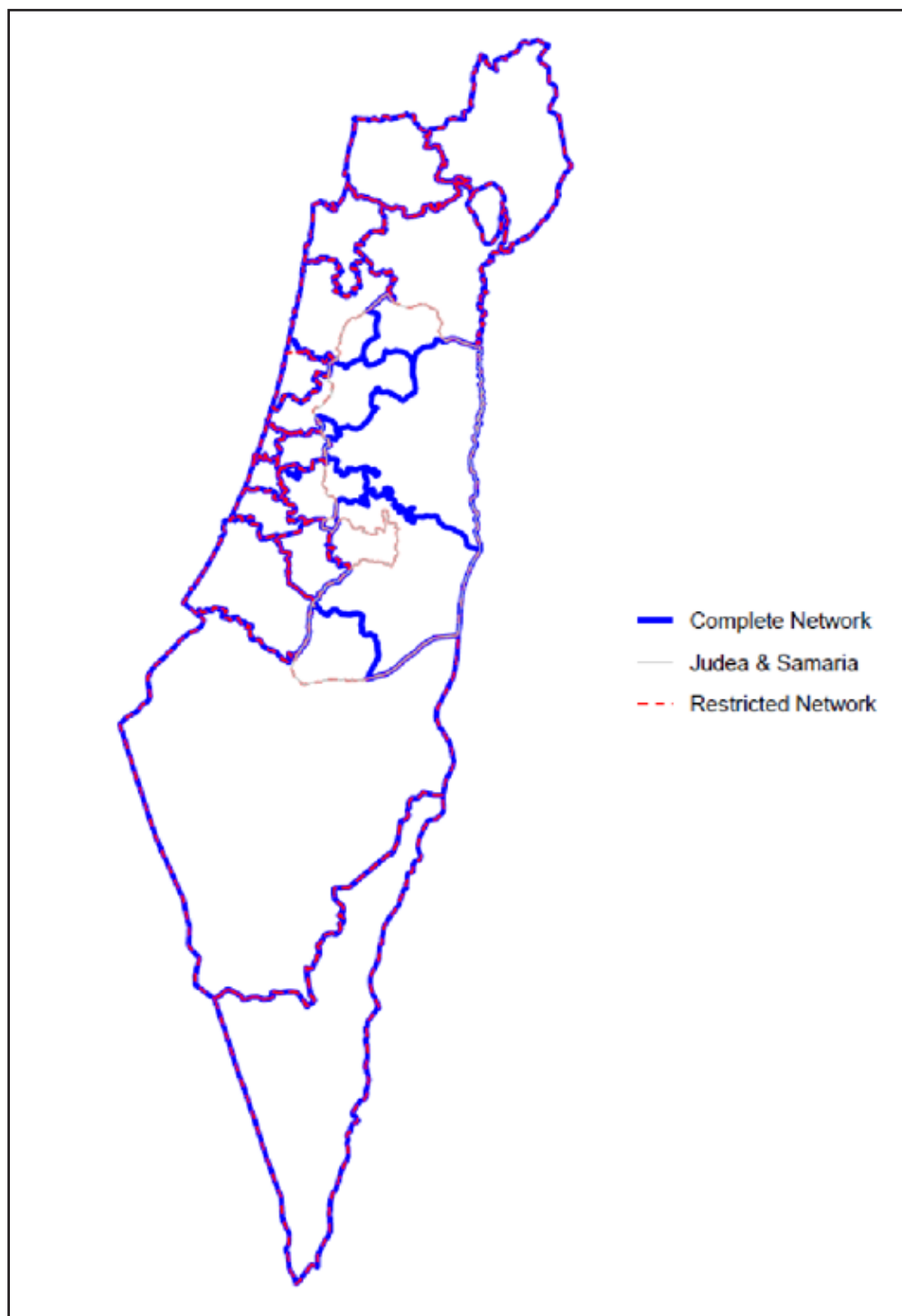
איור נ-1 מציג תוצאות של מדדים שונים ליציבות החלוקה המתקבלת עבור ערכים שונים של  $\gamma$ : מספר הקהילות, ה-*ActualShare* וה-*ExpectedShare*, ומדד Jaccard<sup>3</sup>, שהוא מדד פורמלי למידת הדמיון בין חלוקות שונות (De Bellefon et al., 2021). הקווים האנכיים המקווקווים מציגים את ערכי הפרמטר שנבחר, מכיוון שהוא נמצאים בתוך טווח רחב יחסית בו המדדים שנבחנו מצביעים על חלוקה יציבה.

איור נ-1: מדדי יציבות לתוצאות אלגוריתם זיהוי הקהילות עבור ערכים שונים של  $\gamma$ .  
מקור: עיבודי המחברים



## נספח ב' - אזורי היוממות לפי הרשת המלאה והמוגבלת

מקור: עיבוד המחברים



## נספח ג' - הגדרת אזורי יוממות לפי שיטת TS

הגדרת אזורי יוממות (CZs או Commuting Zones) פותחה בארה"ב בשנות ה-80 במטרה להגדיר כלכלות מקומיות המבוססות על תנועות עובדים בין אזורים. הגדרה זו מבוססת על ההבנה כי החלוקות הגיאוגרפיות המינהליות אינן משקפות פעילות כלכליות אזוריות ואת הגודל של שוק העבודה המקומי ולכן אינן תופסות את כל השונות בשווקי העבודה המקומיים.<sup>4</sup> להגדרה זו יתרונות הן מבחינת מדיניות והן מבחינה מחקרית, כיוון שהיא מצמצמת הטיה שנגרמת כתוצאה מקורלציה מרחבית שאינה נלקחת בחשבון כאשר מנתחים את שוק העבודה תחת יחידות ג"ג קטנות יותר (Carpenter et al., 2022).

אזורי היוםמות מוגדרים באמצעות ניתוח אשכולות היררכי (hierarchical cluster analysis) המבוסס נתוני תנועות עובדים למקום העבודה (Journey-to-Work) הנאספים על ידי לשכת מפקד האוכלוסין של ארה"ב (US Census Bureau). המתודולוגיה לאזורי היוםמות פותחה על ידי Tolbert & Sizer (1987, 1996) (להלן TS). בשלב הראשון, הם הגדירו מטריצת מרחקים  $D$  הקושרת מחוזות על פי תנועות העובדים ביניהם, כאשר כל תא במטריצה מוגדר לפי הנוסחה הבאה:

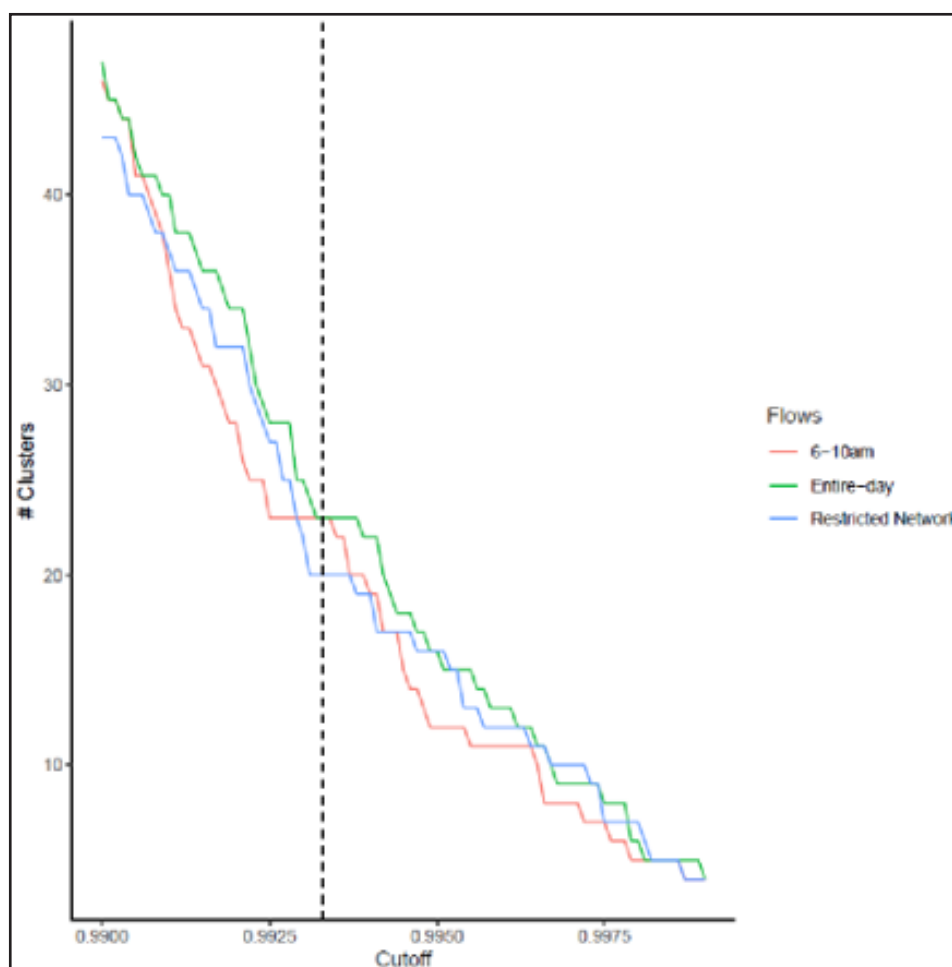
$$D_{ij} = 1 - P_{ij} = 1 - \frac{f_{ij} + f_{ji}}{\min(rlf_i, rlf_j)}$$

המשתנה  $f_{ij}$  מייצג את סך העובדים שנוסעים ממקום מגורים  $i$  למקום עבודה  $j$ , ו- $rlf_i = \sum_j f_{ij}$  מייצג את כוח העבודה המקומי במחוז  $i$ . המשתנה  $D_{ij}$  מגדיר את ה"מרחק" בין שני אזורים, כאשר המרחק אינו מרחק גיאוגרפי אלא מרחק המבוסס על קשרי יוממות בין אזורים. אזורים בעלי קשרי יוממות חזקים הם בעלי "מרחק" נמוך ולהפך.

בשלב השני מטריצת המרחק מוזנת לאלגוריתם של ניתוח אשכולות היררכי. האלגוריתם מצמיד לזוגות את האזורים בעלי המרחק הנמוך ביותר. לאחר מכן כל זוג מוצמד לזוג הקרוב ביותר אליו לפי ממוצע המרחקים בין הזוגות וכך הלאה, עד שכל האזורים מקושרים אחד לשני. החלוקה לאזורי יוממות מתקבלת באמצעות עצירת האלגוריתם ברף שנקבע על בסיס שיקול דעת החוקרים. בפועל, הרף שנקבע ב-TS עומד על 0.98. כלומר, קבוצה של אזורים שהמרחק ביניהם קטן מ-0.98 סווגו כאזור יוממות נפרד.

על מנת לייצר את אזורי היוםמות לפי גישה זו השתמשנו בנתוני הסקר הסלולרי באמצעות הרשת המלאה. כאמור, גם כן יש חשיבות לבחירת הפרמטר שעוצר את התהליך. Tolbert & Sizer (1996) לא סיפקו מתודולוגיה סדורה לבחירת הפרמטר ובפועל הפרמטר נבחר באופן שרירותי על מנת לאזן בין מידת ההתאמה של המחוזות לאזור היוםמות, מחד, לבין הצורך ליצור אזורי יוממות גדולים מספיק על מנת שישקפו שווקי עבודה מקומיים מאידך (Folwer et al., 2016). על מנת לבחור את הפרמטר, נקטנו בגישה דומה לגישה שננקטה בטקסט המרכזי, וחיפשנו אזורים שבהם החלוקה יציבה באופן יחסי. איור נ-2 מציג את מספר אזורי היוםמות לפי גובה הרף. הערך 0.9933 נמצא באזור יציב באופן, הן ברשת המוגבלת, הן ברשת המלאה, והן ברשת הכוללת יוממות בשעות 10-6 בבוקר. ערך זה מייצר מספר אזורים דומה למספר אזורי היוםמות כפי שהתקבל מאלגוריתם זיהוי הקהילות ולפיכך בחרנו בערך זה לצורך השוואה בין החלוקות.

איור נ-2: מספר אזורי היוממות לפי גובה הרף. מקור: עיבודי המחברים



התוצאה הראשונית של האלגוריתם מייצרת אזורי יוממות קטנים ולא רציפים (איור 6 לעיל, בטקסט המרכזי). על מנת לקבל אזורים רציפים הצמדנו את האזורים המנותקים מאזור היוממות שלהם לאזור היוממות הסמוך אליהם. התוצאה הסופית מוצגת באיור 5 לעיל, בטקסט המרכזי.

## הערות:

1. דרגה של צומת היא סכום הנסיעות היוצאות והנכנסות אליה. מכיוון שכל נסיעה תורמת לדרגה של שני צמתים, המכנה כאן כפול מהמכנה של שיעור הנסיעות בפועל, שהוגדר על בסיס סכימת נסיעות ולא סכימת דרגות.
2. כאשר ערך פרמטר הרזולוציה שונה מ-1, המדד עדיין תחום מלמעלה ב-1, אך הערך המצביע על החלוקה האקראית שווה לאחד פחות פרמטר הרזולוציה כפול השיעור הצפוי.
3. מדד Jaccard הוא מדד המאפשר למדוד את מידת הדמיון בין שתי חלוקות אלטרנטיביות באופן פורמלי. בפרט, המדד מוגדר כ:

$$Jaccard \equiv \frac{\sum_i \sum_j I(c_i^1 = c_j^1 \& c_i^2 = c_j^2)}{\sum_i \sum_j I(c_i^1 = c_j^1 \text{ or } c_i^2 = c_j^2)}$$

- כאשר  $c_i^d$  היא הקהילה אליה שויך האזור  $i$  בחלוקה  $(dl, 2)$ ,  $I-1$  ו- $J$  הן פונקציות המקבלות את הערך 1 במידה והתנאי הרשום בתוכן מתקיים, ואת הערך 0 בכל מקרה אחר. מכאן, מדד Jaccard הוא היחס בין סך זוגות האזורים המשוויכים לאותה קהילה בשתי החלוקות לסך זוגות האזורים המשוויכים לאותה קהילה לפחות באחת מהחלוקות. המדד נע בין 0 ל-1, כאשר הערך 0 מתקבל במידה ואין הסכמה בין שתי החלוקות על אף זוג אזורים, ל-1 כאשר החלוקות זהות.
4. <https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/documentation.aspx>

**גל עמדי** הוא חוקר בחטיבת המחקר בבנק ישראל. בוגר תואר ראשון בתוכנית פכ"מ ותואר שני בתוכנית המשולבת לכלכלה ומדיניות ציבורית, שניהם מהאוניברסיטה העברית בירושלים; דוקטורנט לכלכלה באוניברסיטה. תחומי המחקר העיקריים שלו הם כלכלה עירונית וכלכלת תחבורה. [gal.amed@mail.huji.ac.il](mailto:gal.amed@mail.huji.ac.il)

**נדב פורת הירש** הוא חוקר בחטיבת המחקר בבנק ישראל. בוגר תואר ראשון במשפטים וכלכלה ותואר שני בכלכלה, שניהם מהאוניברסיטה העברית בירושלים; דוקטורנט לכלכלה באוניברסיטה. תחומי המחקר העיקריים שלו הם כלכלת עבודה והמגזר העסקי. [nadav.porat@mail.huji.ac.il](mailto:nadav.porat@mail.huji.ac.il)