

# רשתות פסיכולוגיות וקוגניטיביות

## פרויקט- חלק ג'

אביב זילברשטיין 323080556, רון דגני 318170917, אילי ששון 322995358

לפרויקט הגמר שלנו בקורס, הנושא שנבחר הוא נישואין וגירושין. נישואים מייצגים מחויבות לשותפות, אהבה, תקשורת, פתיחות והערכה. בשל המחויבויות העקרוניות הרבים שעליהם עומדים הנישואים, תופעת הגירושין, פירוק נישואין חוקי, הפכה יותר ויותר נפוצה בחברות מודרניות. הבנת הגורמים התורמים להצלחה ולכישלון הזוגי הינה חיונית לזוגות ולאנשי מקצוע שעבודתם בהקשר מערכות יחסים ואף לאינדיבידואלים שמנסים לדייק את תחום זה בחייהם. מטרתנו להבין את הגורמים הללו על ידי מידול תשובותיהם של זוגות נשואים וגרושים על שאלות הנוגעות למערכת היחסים שלהם כרשת תוך הבנת הקשרים בין השאלות וניסיון לאפיין את אופן המענה של גרושים ונשואים. בפרויקט זה, ננסה לצמצם את סט השאלות מ-54 לסט קטן ככל הניתן. נרצה לעשות זאת מכמה סיבות: ראשית, בכך שנצמצם את השאלות, נוכל לזקק יותר טוב ולדעת בבהירות יותר את הסיבות להצלחת קשר זוגי ואת המאפיינים הקריטיים אשר מבדילים בין הגרושים לנשואים עקב הורדת היתירות בשאלות. שנית, מענה על מספר רב של שאלות פוגעת באיכות המידע עקב התעייפות המשתתף והירידה בריכוזו לאורך זמן. לכן, על מנת לעבוד עם הדאטא האיכותי ביותר, נרצה לקצר את זמן המענה להיות הקצר ככל הניתן. פן נוסף אשר צמצום השאלות יכול לעזור בו הינו תקציב: פסיכולוג אשר יושב ושואל זוגות (אשר יתכן כי מקבלים תשלום אשר כולל שקלול של זמן המענה על השאלות) יעלה ככה"נ פחות לאחוז נמוך יותר באופן משמעותי של שאלות. כל אלה יכולים להוביל לתובנות חשובות בנושא וכן לעזור למחקרים עתידיים בתחום להיות יעילים וממוקדים יותר. בנוסף לצמצום השאלות, נרצה לענות על שאלת המחקר שלנו, שהיא האם יש הבדלים פסיכולוגיים מובהקים בין זוגות נשואים לזוגות גרושים ?

## סט הנתונים

לבניית הרשת שלנו השתמשנו במערך נתונים המכיל נתונים מתוך שאלון שעליו ענו כ-170 זוגות המבוסס על שאלות מתוך הטיפול הזוגי של גוטמן. מערך הנתונים נמצא ב-Kaggle והועלה לשם על ידי החוקרים שערכו את הניסוי במטרה ליצור מודל למידה המסווג על סמך התשובות בשאלון האם בני הזוג נשואים או גרושים. הזוגות הם מאזורים שונים בטורקיה שבהם הרשומות נאספו מראיונות פרונטליים של זוגות שכבר היו גרושים או נשואים. למערך הנתונים יש 55 עמודות בסך הכל: 54 מהן מייצגות את 54 השאלות שהזוגות נשאלו, למשל, העמודה הראשונה מכילה את התשובות לשאלה: "אם אחד מאיתנו מתנצל כשהדיון שלנו מתדרדר, הדיון מסתיים". כל התגובות

נאספו בסולם של 5 נקודות (0=אף פעם, 1=לעיתים רחוקות, 2=בממוצע, 3=בתדירות גבוהה, 4=תמיד). העמודה ה-55 מייצגת אם בני הזוג גרושים או נשואים (0=נשואים, 1=גרושים). מכיוון שמתוך 170 הזוגות 86 מהם זוגות גרושים ו-84 מהם זוגות נשואים, אין imbalance בדאטא שלנו כי הדאטא מחולק כמעט במחציתו לפי הלייבל.

## עריכת ועיבוד הדאטא

ראשית נטען את הדאטא ל dataframe של pandas על מנת להקל על הניתוח.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	...	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Divorce
0	2	2	4	1	0	0	0	0	0	0	...	2	1	3	3	3	2	3	2	1	1
1	4	4	4	4	4	0	0	4	4	4	...	2	2	3	4	4	4	4	2	2	1
2	2	2	2	2	1	3	2	1	1	2	...	3	2	3	1	1	1	2	2	2	1
3	3	2	3	2	3	3	3	3	3	3	...	2	2	3	3	3	3	2	2	2	1
4	2	2	1	1	1	1	0	0	0	0	...	2	1	2	3	2	2	2	1	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
165	0	0	0	0	0	0	0	0	0	0	...	1	0	4	1	1	4	2	2	2	0
166	0	0	0	0	0	0	0	0	0	0	...	4	1	2	2	2	2	3	2	2	0
167	1	1	0	0	0	0	0	0	0	1	...	3	0	2	0	1	1	3	0	0	0
168	0	0	0	0	0	0	0	0	0	0	...	3	3	2	2	3	2	4	3	1	0
169	0	0	0	0	0	0	0	1	0	0	...	3	4	4	0	1	3	3	3	1	0

170 rows × 55 columns

נציין כי אין ערכים חסרים בדאטא ולכן לא היה צורך לטפל באספקט זה בדאטא. לאחר מכן על מנת ליצור גרף לכל קבוצה, נסנן את הטבלה לפי הערכים הרלוונטיים לכל קבוצה ונקבל טבלה של הקבוצה בלבד. על מנת ליצור גרף של הקבוצה נבצע את השלבים הבאים:

נחשב מטריצת קורלציות ע"י שימוש בקורלציית פירסון בין ערכי העמודות (אם נרצה קורלציה בין האנשים ולא בין השאלות נעשה היפוך transpose של המטריצה, משום שאנו בודקים קורלציה בין התשובות לשאלות שאלו כרגע העמודות). במקרים בהם סט השאלות מצומצם יכולים להיווצר מצבים שבהם קיימים זוגות שענו בצורה זהה על כל השאלות, במצב זה סטיות התקן אפסיות וערך המתאם לא מגודר ע"י נוסחת פירסון:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

כדי להתמודד עם מצב זה, מילאנו את הערכים הריקים במטריצת הקורלציה ברעש גאוסיאני סטנדרטי התחום בין [-1,1]. יש לציין כי מדובר על מקרים בודדים בלבד ולכן אינם מהווים הטיה משמעותית לדאטא.

נקבל מטריצת הקורלציות (ריבועית כמובן) אשר תהיה בגודל מספר העמודות X מספר העמודות, להלן צילום חלקי של המטריצה:

	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Q1	1.000000	0.819066	0.832508	0.825066	0.881272	0.287140	0.427989
Q2	0.819066	1.000000	0.805876	0.791313	0.819360	0.102843	0.417616
Q3	0.832508	0.805876	1.000000	0.806709	0.800774	0.263032	0.464071
Q4	0.825066	0.791313	0.806709	1.000000	0.818472	0.185963	0.474806
Q5	0.881272	0.819360	0.800774	0.818472	1.000000	0.297834	0.381378
Q6	0.287140	0.102843	0.263032	0.185963	0.297834	1.000000	0.424212
Q7	0.427989	0.417616	0.464071	0.474806	0.381378	0.424212	1.000000

על מטריצה זו נבצע פלטור ע"י שימוש ב- tmfg שיטה לבחירת הקשרים המשמעותיים ביותר (פירוט מלא במאמר המצורף בנספחים). לאחר הסינון, נתייחס למשקלי הקשתות כבינאריים- אלו שעברו את הסינון יקבלו משקל 1 ואלו שלא 0.

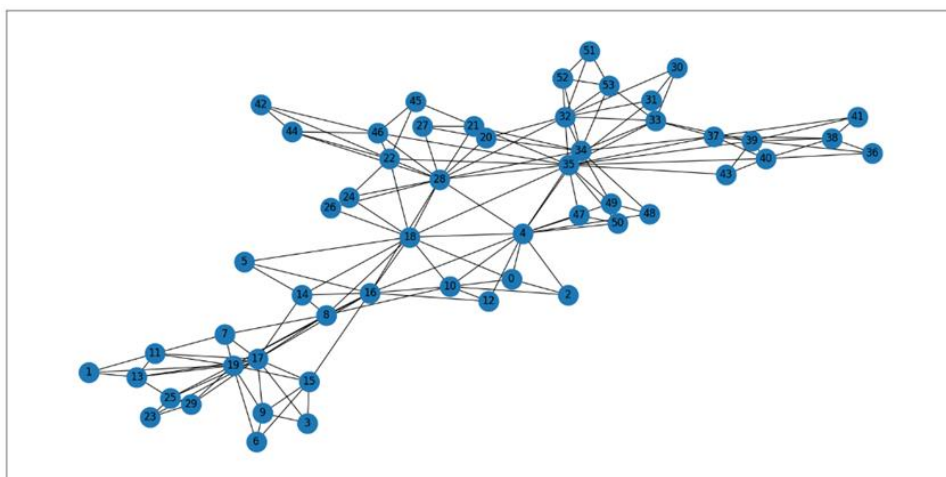
לבסוף, נוכל להדפיס את הגרף הכולל את הקשתות שנבחרו ואת הצמתים שהוגדרו על ידינו מראש כעמודות הטבלה (שאלות\זוגות).

ברשתות שלנו אין תכונות נוספות על הצמתים או על הקשתות, ואין משקלים על הקשתות. הסיבה לכך היא שאנחנו רוצים להתייחס אל ההקשר בין השאלות יותר מאשר להתייחס לכמות הדמיון או אי-הדמיון בין התשובות של הזוגות. נציין כי השאלות והזוגות ממסופרים וכך נקבע הייצוג שלהם אשר מוצג בגרף.

להלן הנתונים הספציפיים על כל רשת לחלק א' של הפרויקט:

רשת הכוללת את כל הדאטא:

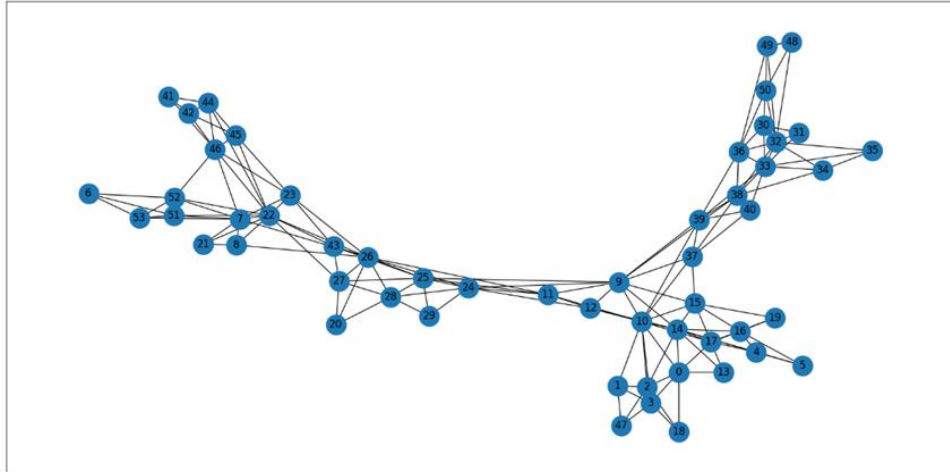
nodes: 54  
edges: 156



הקודקודים מייצגים שאלות.

רשת הכוללת את הזוגות הנשואים בלבד:

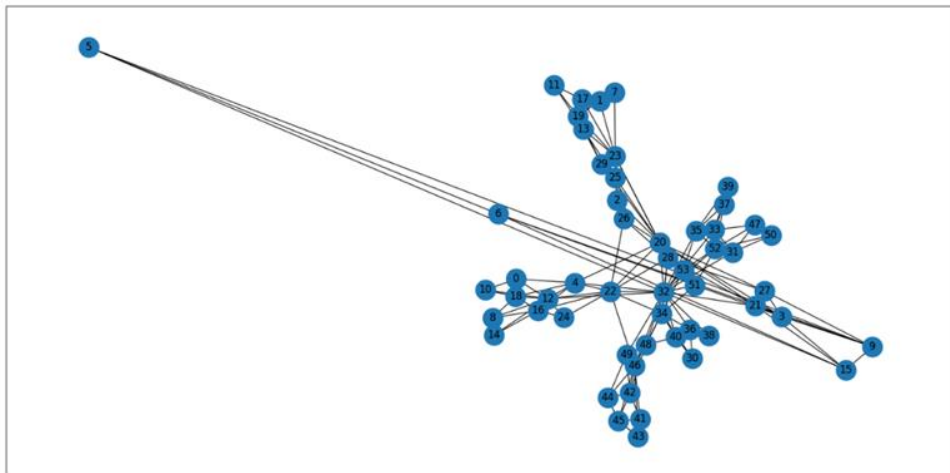
nodes : 54  
edges : 156



הקודקודים מייצגים שאלות .

רשת הכוללת את הגרושים בלבד:

nodes : 54  
edges : 156



הקודקודים מייצגים שאלות .

כמו כן, בבנה רשתות הבוחנות את המדדים על קהילות פרטניות מתוך אילו שנמצאו בניתוחים הקודמים.

## ויזואליזציה

בתהליך הוויזואליזציה התמקדנו במספר דברים על מנת לייצר ויזואליזציה איכותית. ראשית, הגדלנו את גודל הגרף במידה רבה על מנת לקבל תמונה טובה ולמנוע עומס על המשתמש. הגרף הינו בינארי, לכן כל הקשתות הינן באותו העובי. שיפור נוסף של הקריאות התבצע על ידי `spring_layout` לתצוגה טובה יותר של הקהילות ושינוי  $K$  שמייצג את המרחק בין הצמתים. לכל גרף עשינו התאמה של  $k$  לתצוגה הטובה ביותר. בכל הגרפים למעט שניים שעליהם נרחיב בהמשך, השתמשנו בנוסחה:

$$\frac{1}{\# nodes_{multi_k=1}} = \frac{multi_k}{\# nodes}$$

בניגוד לערך הדיפולטיבי של  $K$  שהוא 1 חלקי שורש מספר הצמתים. לגרף הגרושים בגלל צומת 5 הרחוקה השתמשנו בערך הדיפולטיבי כדי לקבל את הגרף הקריא ביותר. בגרף כל הזוגות כשהצמתים הם הזוגות בגלל הכמות הגדול של הצמתים שיוצר עומס, הגדרנו את  $multi_k$  להיות שווה ל-30 כדי להקל על הקריאה ולרווח את הגרף.

שנית, בבחירת הצבעים בחרנו ב `cmap` אשר לא תקשה על עיוור צבעים ותקל על זיהוי הקהילות השונות. ה-`cmap` שנבחרה היא `set2`.

## ניתוח ומחקר

שאלת המחקר שלנו היא: האם יש הבדלים פסיכולוגיים מובהקים בין זוגות נשואים לזוגות גרושים?

השערת המחקר שלנו היא שקיימים הבדלים פסיכולוגיים בין זוגות גרושים לזוגות נשואים המתבטאים באופן המענה שלהם לשאלות. השאלות בוחנות את הקשר ואנו סבורים כי השוני באופי הקשר והצלחתו יגרום להבדל במענה על השאלות.

## חלק א:

כדי לענות על שאלת המחקר שלנו בחרנו למצוא ברשתות קהילות של שאלות שיש ביניהן קשרים קורלטיביים. המטרה של מציאת הקהילות בכל רשת היא ליצור קבוצות של שאלות שיש קורלציה גבוהה בתשובות של זוגות גרושים/נשואים/שניהם עליהן. את הקהילות ניצור על ידי שימוש באלגוריתם Louvain אשר נלמד בהרצאה ועשינו בו שימוש בתרגיל בית 2. האלגוריתם של Louvain ממקסם באופן איטרטיבי את מדד המודולריות, מה שמוביל לזיהוי קהילות בקנה מידה שונה וחושף מבני קהילה מקוננים או חופפים בתוך הרשת.

היתרון למידול הדאטא ברשת, הינו שבאמצעות הרשת, ניתן לראות את המבנה הטופולוגי של הקשר בין השאלות\הזוגות לפי הקורלציות שלהם ולבחון אותם בעזרת מדדי רשת ומרכזיות שעל בסיסם יהיה ניתן לקבל תובנות על התופעות בדאטא. יתרון נוסף במידול דאטא זה כרשת הינו האפשרות להפוך את התפקידים של הקודקודים מתפקיד של שאלה לתפקיד של זוג ובך לקבל מידע לגבי טיב ההפרדה ואופן הקשר (והמדדים) שיש בין הזוגות ולא רק בין השאלות, בקלות מאוד. החלפת הצמתים מהשאלות לזוגות יאפשר לנו לבחון את השאלות במרכז תחילה, ולהבין את הדינמיקה בין השאלות ואיזה שאלות הן המשמעותיות ביותר ולאחר מכן להסתכל על הזוגות כאשר הם עונים על שאלות אלו.

נסתכל על טבלת הערכים ותוך כדי הרחבה על כל מדד נתייחס למשמעות הערכים:

	divorced	married	both married and divorced
ACC	0.695191	0.690423	0.728416
Modularity of Louvain Communities	0.618179	0.626705	0.554610
ASPL	2.983229	3.691824	2.929420
Average Group Betweenness Centrality	0.109033	0.138601	0.149760
Average Degree	5.777778	5.777778	5.777778
Average Maximal Number of Cliques per Node	51.000000	51.000000	51.000000
Average Connectivity	3.104822	3.104822	3.104822
Average Degree Centrality	0.109015	0.109015	0.109015
Average Group Closeness Centrality	0.438760	0.348124	0.438563
Average Eigenvector Centrality	0.109909	0.115602	0.111869

בניתוח בין הרשתות שלנו השתמשנו בעשרת המדדים הבאים :

- **Average Clustering Coefficient (ACC) - הממד מייצג את הסבירות ששני שכנים של**

צומת מחוברים זה לזה בממוצע על כל הצמתים ברשת. ערך הקרוב ל-1 מצביע על כך שהצמתים ברשת נוטים ליצור אשכולות הקשורים זה לזה, דבר המצביע על רמה גבוהה של מקבץ מקומי ונוכחות של קהילות הדוקות בתוך הרשת. במחקרנו, רשת עם ערך גבוה בממד זה, מעיד על רשת שבה מספר הקשרים הקורלטיביים המובהקים בין שאלות על פי התשובות של הזוגות גדול יותר, מכיוון שזוגות רבים עונים בצורה דומה על מספר שאלות רב יותר. ה ACC הגבוה ביותר נמדד עבור רשת כל הזוגות (0.728) מה שמצביע שהאשכולות ברשת זו הדוקים יותר והרשת צפופה יותר בהשוואה לרשת הזוגות הנשואים ורשת הזוגות הגרושים. תוצאה זו הייתה צפויה מראש, שכן רשת כל הזוגות לוקחת בחשבון את כל הזוגות במאגר הנתונים, ולכן סביר שיימצאו עוד קשרים קורלטיביים מובהקים בין השאלות. ה ACC של רשת הזוגות הגרושים (0.695) היה גדול מזה של רשת

- הזוגות הנשואים (0.69), מה שיכול להצביע שנמצאו יותר קשרים קורלטיביים מובהקים היטב בין השאלות על סמך התשובות שלהם. מכך נוכל להסיק שהשונות של תשובות הזוגות הגרושים נמוכה יותר מזו של תשובות הזוגות הנשואים. יש לציין שעקב ההפרש הקטן בין ערכי ה-ACC מסקנה זו אינה מובהקת ממדד זה בלבד.
- מודולריות על פי הקבוצות של Louvain** - המדד מייצג את המידה שבה רשת מחולקת לקהילות או מודולים נפרדים ומקושרים זה בזה. ערך קרוב ל-1 עבור מודולריות מצביע על רמה גבוהה של מבנה קהילתי בתוך רשת, כלומר הרשת מחולקת באופן משמעותי לקהילות נפרדות ממה שניתן היה לצפות ברשת אקראית. במחקרנו, רשת עם ערך גבוה במדד זה, בהינתן הקהילות שנוצרו באמצעות אלגוריתם Louvain מעיד שהן משמעותיות, וכי יש קבוצות מוגדרות היטב של שאלות שיש ביניהן קשרים קורלטיביים מובהקים. המודולריות של רשת כל הזוגות הייתה הנמוכה ביותר (0.554), בעוד שערכי המודולריות של רשת הזוגות הגרושים (0.618) ורשת הזוגות הנשואים (0.626) היו גבוהים יותר. תוצאה זו הייתה גם כן צפויה מראש, משום שהורדת התשובות של חצי מהזוגות כחלק מבניית רשתות הגרושים ורשתות הנשואים גרמה לרשתות להיות עם פחות קשתות בין שאלות שונות מה שמקל על חלוקתן לקהילות של שאלות שונות ומוגדרות היטב. המודולריות של הקהילות ברשת הזוגות הנשואים הייתה גדולה יותר מזו של הקהילות ברשת הזוגות הגרושים, מה שעלול להצביע על כך שהקהילות ברשת הנשואים משמעותיות יותר מאילו של הרשת השנייה. יש לציין שעקב ההפרש הקטן בין ערכי המודולריות מסקנה זו אינה מובהקת ממדד זה בלבד.
  - Average Shortest Path Length (ASPL)** - מדד זה מייצג את המספר הממוצע של שלבים או קצוות הנדרשים למעבר בין שני צמתים כלשהם ברשת. ערך גבוה עבור אורך הנתיב הממוצע הקצר ביותר מצביע על כך שהצמתים ברשת מחוברים מאוד, עם מרחקים קצרים יחסית ביניהם, דבר המצביע נגישות בתוך הרשת. במחקרנו, רשת עם ערך נמוך במדד זה מעיד על רשת שמספר השאלות שעבורן קיימים קשרים קורלטיביים מובהקים היטב גדול אשר נמצאים באזורים נרחבים ברשת. ה-ASPL הגבוה ביותר נמדד על רשת הזוגות הנשואים (3.691) בהשוואה אצל רשת הגרושים (2.983) ורשת כל הזוגות (2.929).
  - Average Group Betweenness Centrality (AGBC)** - מדד זה מייצג את המידה שבה קבוצה של צמתים פועלת באופן קולקטיבי כגשרים או מתווכים בחיבור חלקים שונים של הרשת. ערך קרוב ל-1 עבור AGBC ברשת מצביע על כך שקבוצות הצמתים יש השפעה קולקטיבית גבוהה בשליטה על זרימת המידע או המשאבים בין חלקים שונים של הרשת בממוצע. במחקרנו, רשת עם ערך גבוה במדד זה היא רשת שבה סביר שיהיו קבוצות של

שאלות אשר מתווכות בין מספר קהילות של שאלות באזורים שונים ברשת אשר יכולות להעיד על שאלות שיכולות לייצג מספר אספקטים שונים במקביל של זוגיות אשר נמדדו בשאלון כגון תקשורת, עבודת צוות, סובלנות ועוד. את המדד הזה חישבנו על כל קהילה שנוצרה על ידי האלגוריתם של Louvain בנפרד (שכן ככה הוא ממומש ב-network) ועשינו ממוצע. ערך ה-AGBC הגבוה ביותר נצפה ברשת כל הזוגות (0.158), הזוגות הנשואים (0.138) והערך הנמוך ביותר נצפה ברשת הגרושים (0.109).

- **Average Degree** - מדד זה מייצג את המספר הממוצע של חיבורים או קצוות שיש לכל צומת ברשת. ערך גבוה של המדד הממוצע מצביע על כך שבממוצע, לצמתים ברשת יש מספר רב של חיבורים, דבר המצביע על רשת שאלות שביניהן יש קשרים קורלטיביים מובהקים רבים. במחקרנו, רשת עם ערך גבוה במדד זה, מעיד על רשת שבה מספר הקשרים הקורלטיביים המובהקים בין שאלות על פי התשובות של הזוגות גדול יותר, מכיוון שזוגות רבים עונים בצורה דומה על מספר שאלות רב יותר. באופן מפתיע, הדרגה הממוצעת של כל הרשתות שחקרנו בחלק זה הייתה זהה (5.778). כלומר, בהינתן התשובות של זוגות גרושים/ זוגות נשואים/ כל הזוגות על השאלון, קיימות שאלות אשר מחוברות על ידי קשרים קורלטיביים מובהקים אל בממוצע 5.7 שאלות אחרות. יתכן שתוצאה זו נובעת מיצירת הרשתות לפי ה-tmfgr משום שהפילטר גורם לבחירה של מספר קבוע של קשתות ביחס למספר הקודקודים ואז כאשר ממצעים את אלו, המדד שווה בין הרשתות.

- **Average Degree Centrality (ADC)** - מדד זה מייצג את מספר החיבורים הממוצע שיש לכל צומת, ומציין את רמת ההשפעה הכללית או החשיבות של הצמתים על סמך הקישוריות שלהם. ערך גבוה של ה-ADC מצביע על כך שלצמתים ברשת יש מספר גדול יותר של חיבורים בממוצע, מה שמצביע על רמה גבוהה יותר של השפעה או מרכזיות בתוך הרשת. במחקרנו, רשת עם ערך גבוה במדד זה, מעיד על רשת שבה מספר הקשרים הקורלטיביים המובהקים בין שאלות על פי התשובות של הזוגות גדול יותר, מכיוון שזוגות רבים עונים בצורה דומה על מספר שאלות רב יותר. קודקודי שאלות בעלי מרכזיות גבוהה יכולות להיות מרכזיות ברשת. ה-ADC של כל הרשתות שחקרנו בחלק זה הייתה זהה (0.109). כלומר, בהינתן התשובות של זוגות גרושים/ זוגות נשואים/ זוגות נשואים וגם זוגות גרושים על השאלון, ההשפעה הממוצעת של שאלה ברשת היא בערך על 10 אחוז מסך כל השאלות האחרות ברשת. גם כאן, יתכן שתוצאה זו נובעת מיצירת הרשתות לפי ה-tmfgr משום שהפילטר גורם לבחירה של מספר קבוע של קשתות ביחס למספר הקודקודים ואז כאשר ממצעים את אלו, המדד שווה בין הרשתות.



- **Average Group Closeness Centrality (AGCC)** - מדד זה מייצג את הסכום ההפוך הממוצע של מרחקי הנתיב הקצרים ביותר בין כל צומת בקבוצה, המציין באיזו יעילות מידע או משאבים יכולים להתפשט בתוך הקבוצה. ערך גבוה עבור AGCC ברשת מציין שהצמתים בתוך קבוצה קשורים זה בזה, מה שמאפשר תקשורת יעילה והפצה מהירה של מידע או משאבים בתוך הקבוצה, מה שמשפר את שיתוף הפעולה והלכידות. במחקרנו, רשת עם ערך גבוה במדד זה היא רשת שבה סביר שיהיו קבוצות של שאלות אשר מתווכות בין מספר קהילות של שאלות באזורים שונים ברשת אשר יכולות להעיד על שאלות שיכולות לייצג מספר אספקטים שונים במקביל של זוגיות אשר נמדדו בשאלון כגון תקשורת, עבודת צוות, סובלנות ועוד. את המדד הזה חישבנו על כל קהילה שנוצרה על ידי האלגוריתם של Louvain (שכן ככה הוא ממומש ב-networkx) ועשינו ממוצע. ערך ה-AGCC הגבוה ביותר נצפה ברשת הגרושים (0.438), הערך השני הכי גבוה נצפה ברשת כל הזוגות (0.437) והערך הנמוך ביותר נצפה ברשת הנשואים (0.348).
- **Average Connectivity (AC)** - מדד זה מייצג את המספר המינימלי הממוצע של צמתים שיש להסיר על מנת לנתק צומת משאר הרשת. ערך גבוה עבור מדד ה-AC מצביע כי הרשת חזקה וגמישה, שכן היא תדרוש הסרה של מספר משמעותי של צמתים כדי לשבש את הקישוריות בין הצמתים, מה שמרמז על קישוריות חזקה ויתירות במבנה הרשת. במחקרנו, אם נסתכל על הערך עבור כל צומת ברשת, ערך נמוך מעיד על כך שיש מספר שאלות ברשת אשר מספר הקשרים הקורלטיביים המובהקים בינן לבין שאלות אחרות נמוך ולכן הן יתירות ונוכל לא להתייחס אליהן בחלק השני של הפרויקט. לעומת זאת, ערך גבוה מעיד על שאלות מרכזיות ומשמעותיות. ה-AC של כל הרשתות היה זהה ועמד על 3.104. כלומר, בהינתן התשובות של זוגות גרושים/ זוגות נשואים/ זוגות נשואים וגם זוגות גרושים על השאלון, קיימות שאלות אשר צריך לנתק בממוצע לפחות 3 שאלות אחרות על מנת שנוכל לבדד אותן מהרשת. מספר זה קטן מאוד ממספר הצמתים ברשתות שלנו (55) ולכן זה מעיד על רשתות חלשות ושבירות בעלות פוטנציאל גדול לבידוד של שאלות מתוכן שכן מספר הקשרים הקורלטיביים המובהקים בין השאלות אינו גדול מספיק. גם כאן, יתכן שתוצאה זו נובעת מיצירת הרשתות לפי ה-tmfgr משום שהפילטר גורם לבחירה של מספר קבוע של קשתות ביחס למספר הקודקודים ואז כאשר ממצעים את אלו, המדד שווה בין הרשתות.
- **Average Maximal Number of Cliques per Node (AMNCN)** - מדד זה מודד את המספר הממוצע של קליקות נפרדות שכל צומת שייך אליהן, כאשר קליקות הן תת-גרפים מחוברים במלואם בתוך הרשת. ערך גבוה של ה-AMNCN מצביע על כך שצמתים ברשת מעורבים במספר רב של תתי-קבוצות או קהילות שונות, מה שמצביע על רמה גבוהה של

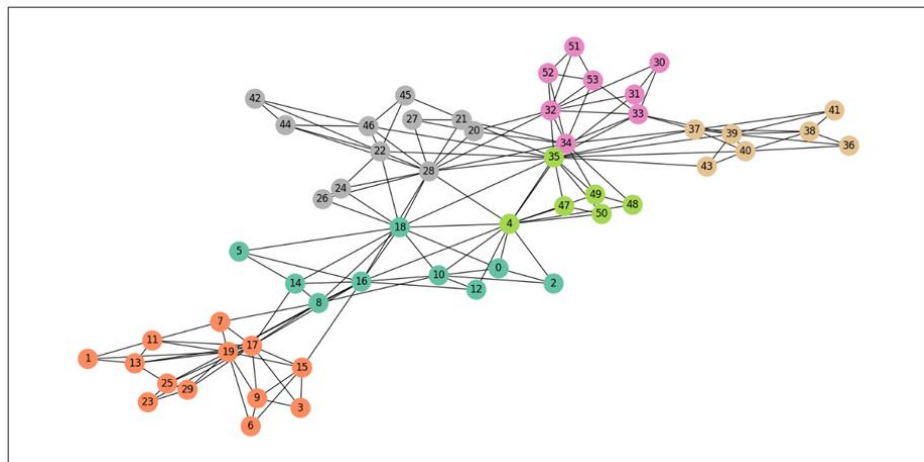
חפיפה של שאלות בתוך הרשת. באופן מפתיע, המספר המרבי הממוצע של קליקות לצומת היה זהה (51). כלומר, בהינתן התשובות של זוגות גרושים/ זוגות נשואים/ זוגות נשואים וגם זוגות גרושים על השאלון, כל שאלה משתייכת בממוצע אל 51 קליקות שונות. תוצאה זו מצביעה על מבנה רשת מפוצל מאוד עם קישוריות מינימליות בין השאלות, מה שתואם את הערכים הנמוכים יחסית שקיבלנו אצל מדדי ה-AGB וה-ADC. כמו המדדים הקודמים, גם במדד זה יתכן כי זה נובע מאופן סינון הקשתות באמצעות t.mfg.

- **Average Eigenvector Centrality (AEC)** - מדד זה מודד את מרכזיות הצמתים בממוצע על כל הרשת בהתבסס על התפיסה שחשיבותו של צומת מושפעת לא רק מהחיבורים המיידיים שלו אלא גם מהחשיבות של הצמתים המחוברים שלו. ערך גבוה במדד זה מצביע על כך שבממוצע על הרשת, לצומת יש קשרים לצמתים מרכזיים אחרים, מה שמצביע על כך החזקת עמדה של השפעה או חשיבות בתוך הרשת. צמתים בעלי eigenvector centrality גבוה מחוברים לצמתים אחרים בעלי מרכזיות גבוהה בעצמם, מה שמוביל לאפקט מחזק המגביר את החשיבות הכוללת שלהם במבנה הרשת. הוספנו את מדד זה בעקבות ההרצאה בנושא, והמלצתו החמה להשתמש במדד זה עקב היציבות שלו. אצלנו, ערך ה-AEC הגבוה ביותר נצפה אצל רשת כל הזוגות (0.1118), הערך השני הכי גבוה נצפה אצל רשת הזוגות הנשואים (0.1156) והערך הכי נמוך נצפה אצל רשת הזוגות הגרושים (0.1099). מכיוון שההפרשים בין המדדים קטנים מאוד, קשה להסיק שיש הבדלים מובהקים בין הרשתות על פי מדד זה בפן הגלובלי.

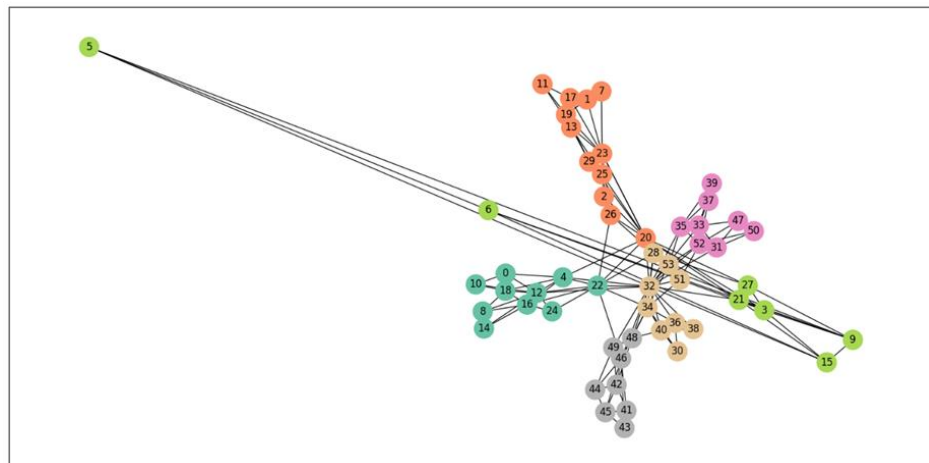
בחירת המדדים הייתה משימה בה היו לנו לבטים רבים, עקב מספר המדדים הרב שיש בתחום ניתוח הרשתות ועקב הצורך בהבנה עמוקה של המשמעות של כל מדד במקרה של הרשת בה הוא נמדד. לבסוף החלטנו לבחור במדדים אלה משום שאלו נראו לנו מקיפים ורלוונטיים ( גם אלו אשר קבועים עקב הפילטור הינם חשובים על מנת לקבל הבנה על פריסת הרשתות).

בעת, נסתכל על הוויזואליזציה של הקהילות לפי אלגוריתם Louvain:

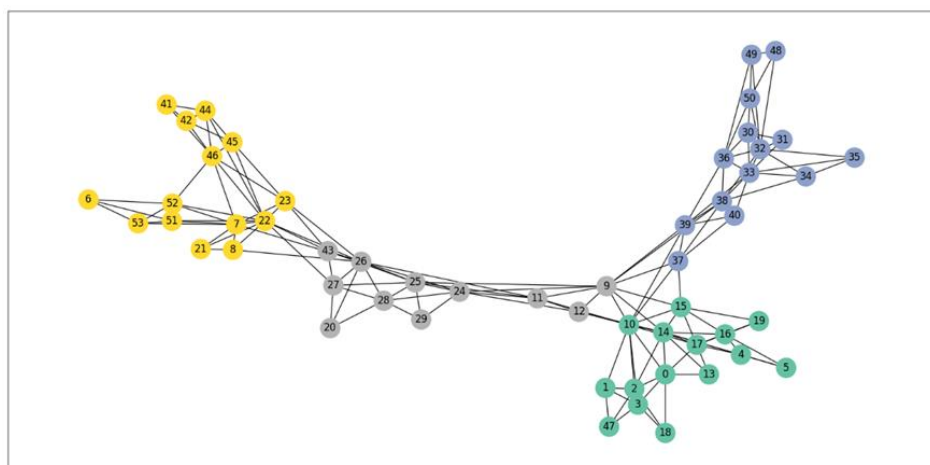
- רשת כל הזוגות:



- רשת הזוגות הגרושים:



- רשת הזוגות הנשואים:



נשים לב כי יש חלוקה ברורה לקהילות לפי האלגוריתם של Louvain. נשווה את החלוקה שהתקבלה על ידי האלגוריתם לחלוקה שלנו לקהילות לפי סיווג פסיכולוגי שלנו ונסתמך על הקהילות בבחירת השאלות המשמעותיות.

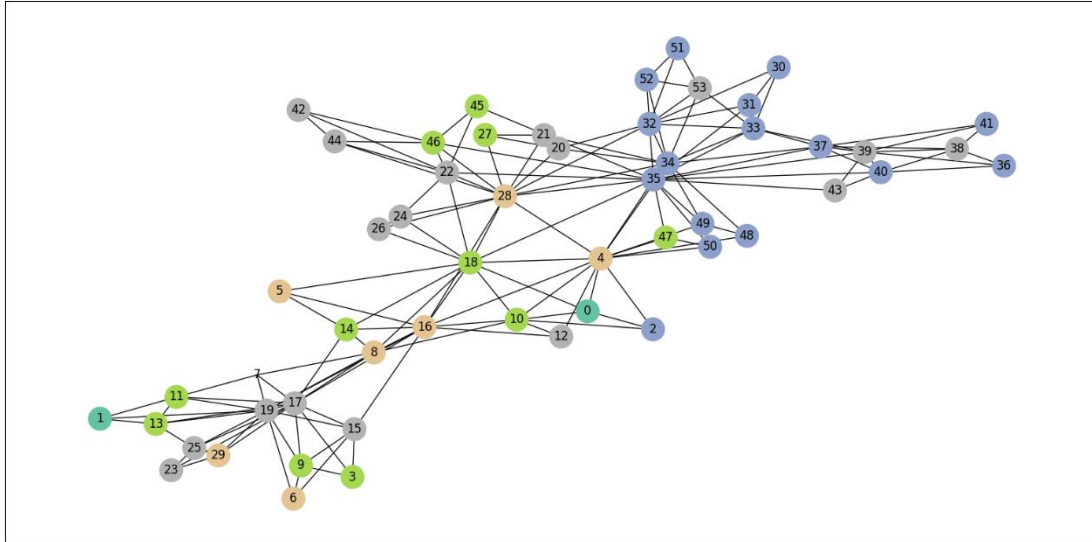
נבחן את הקהילות לפי הסיווג ההגיוני לנו כשאנחנו מנסים לנתח את השאלות מבחינת פסיכולוגית. בחרנו לחלק את השאלות ל-5 קבוצות ( בחרנו ב-5 עקב הסכמה כי ישנם 5 נושאים אליהם השאלות מתקשרות) השאלות חולקו כלהלן:

סלחנות	סובלנות	עבודת צוות	הכרת טוב	תקשורת
1	3	4	5	7
2	31	10	6	13
	32	11	7	16
	33	14	9	18
	34	15	17	20
	37	19	29	21
	38	28	30	22
	41	46		23
	42	47		24
	49	48		25
	50	12		26
	51			27
	52			39
	53			40
	35			43
	36			44
				45
				54

הסיווג נעשה על מנת להסתכל על קבוצות שונות של השאלות ולנסות להבין מה השאלות השונות מנסות לבחון, זאת מהמחשבה שבה בקשר יש מספר של מאפיינים וכל שאלה מנסה לבחון מאפיין

אחר בקשר, ויהיה ניתן לחלק את השאלות לקהילות על ידי התייחסות למשמעות הפסיכולוגית שלהן.

נסתכל על הגרף שבו השאלות הן הצמתים והצבעים נקבעים לפי הסיווג שלנו של השאלות.



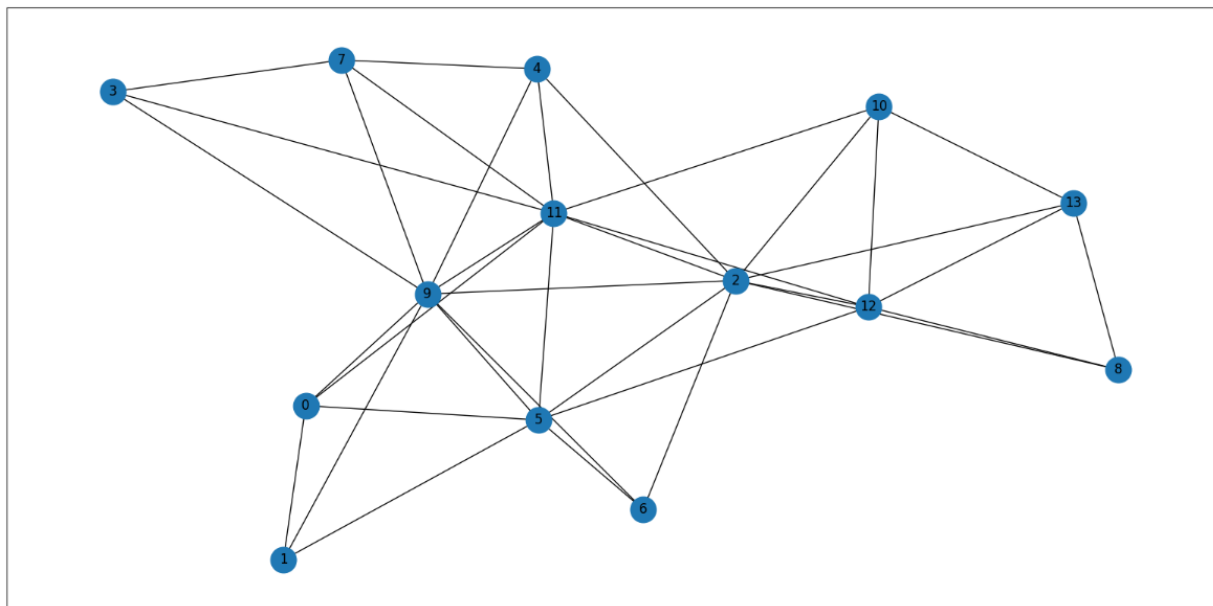
נשים לב כי באופן מפתיע, הקהילות שמצאנו על פי הסיווג הפסיכולוגי של השאלות הן שונות באופן משמעותי לקהילות שנמצאו על ידי אלגוריתם לוויין, וכי לא נראה שהן מסודרות בקהילות מבחינת הגרף. הסבר אפשרי לשנוי הוא שהאלגוריתם מתמקד בגרף, ובעוד שאנחנו הסתכלנו על המשמעות הפסיכולוגית. התוצאה המפתיעה הובילה אותנו למחשבה לדרך שבה יהיה ניתן לבחון את הסוגייה ואת טיב הסיווג שלנו. על מנת לקבל סיווג פסיכולוגי טוב וחד משמעי יותר ניתן לתת לקבוצה של אנשים גדולה יותר או למומחים לסווג את השאלות ובכך לקבל סיווג אנושי אמין יותר מאשר הסיווג שלנו בלבד. אי התאמה של הסיווג תוך הסתכלות על המשמעות פסיכולוגית, גורם לנו לחשוב מחדש על ההנחה, ששאלות שישנן יתירות בניהן יהיו שאלות שחוקרות את אותו דבר בקשר בין בני הזוג. בנוסף, מידת ההסברה (expansibility) של לוביין והקישור של האלגוריתם למשמעות הפסיכולוגית הוא כזה שקשה יהיה להסביר, כי הקהילות לפי לוויין נבחרות מהסתכלות על הגרף כגרף, להסביר כיצד הקהילה נקבעה יהיה בעייתי בשל כך.

כעת, נבחן את הקהילות שנוצרו מתוך אלגוריתם לוביין באופן פרטני, ישנן מספר רב של קהילות ולכן נציג קהילה אחת מכל אחד מהגרפים עם חישובי מדדים עבור כל צומת בקהילה(התהליך בוצע עבור כל אחת מהקהילות):

עבור גרף כל הזוגות:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Degree Centrality	0.307692	0.230769	0.692308	0.230769	0.307692	0.538462	0.230769	0.307692	0.230769	0.692308	0.307692	0.692308	0.461538	0.307692
Eigenvector Centrality	0.217839	0.159071	0.397683	0.167475	0.234140	0.342457	0.189220	0.201090	0.142819	0.388627	0.211436	0.409338	0.281108	0.173175
Betweenness Centrality	0.006410	0.000000	0.234402	0.000000	0.010043	0.094872	0.000000	0.004274	0.000000	0.184615	0.012179	0.204060	0.065385	0.004274

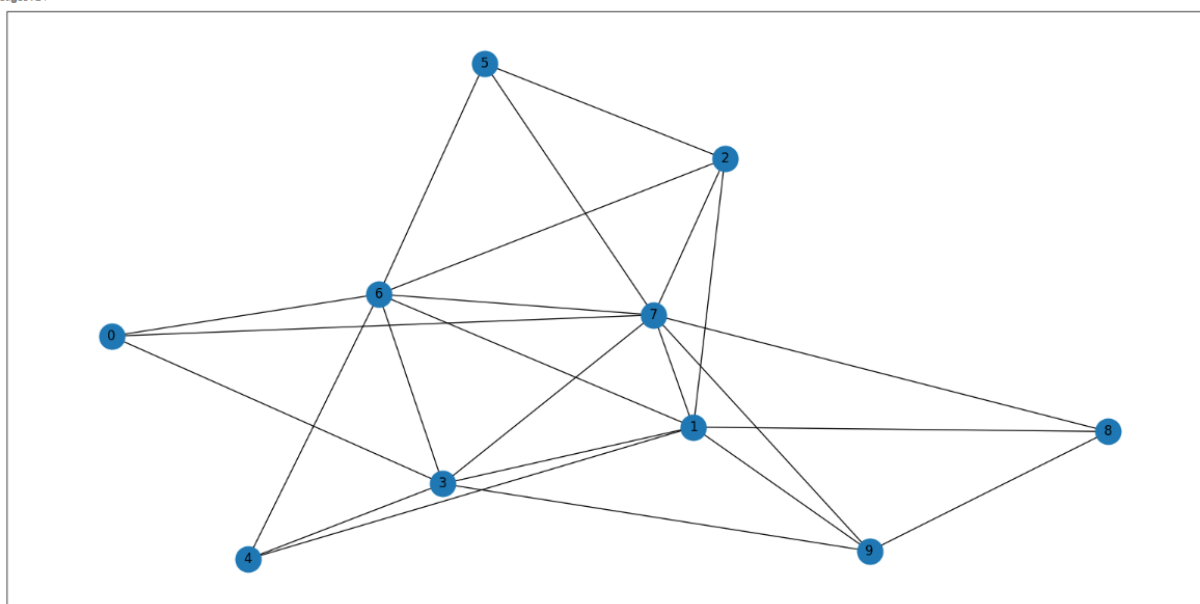
nodes:14  
edges:36



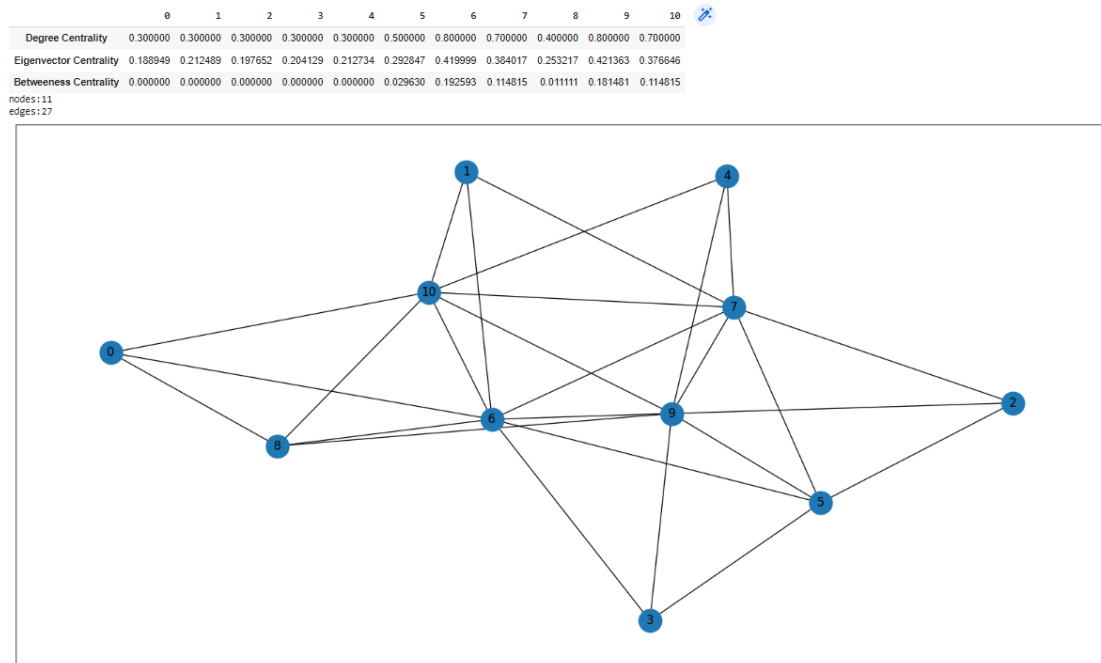
עבור גרף הגרושים:

	0	1	2	3	4	5	6	7	8	9
Degree Centrality	0.333333	0.777778	0.444444	0.666667	0.333333	0.333333	0.777778	0.888889	0.333333	0.444444
Eigenvector Centrality	0.227606	0.409759	0.274439	0.369303	0.221229	0.209858	0.403383	0.443840	0.209858	0.268062
Betweenness Centrality	0.000000	0.134259	0.009259	0.069444	0.000000	0.000000	0.134259	0.226852	0.000000	0.009259

nodes:10  
edges:24



## עבור גרף הנשואים:



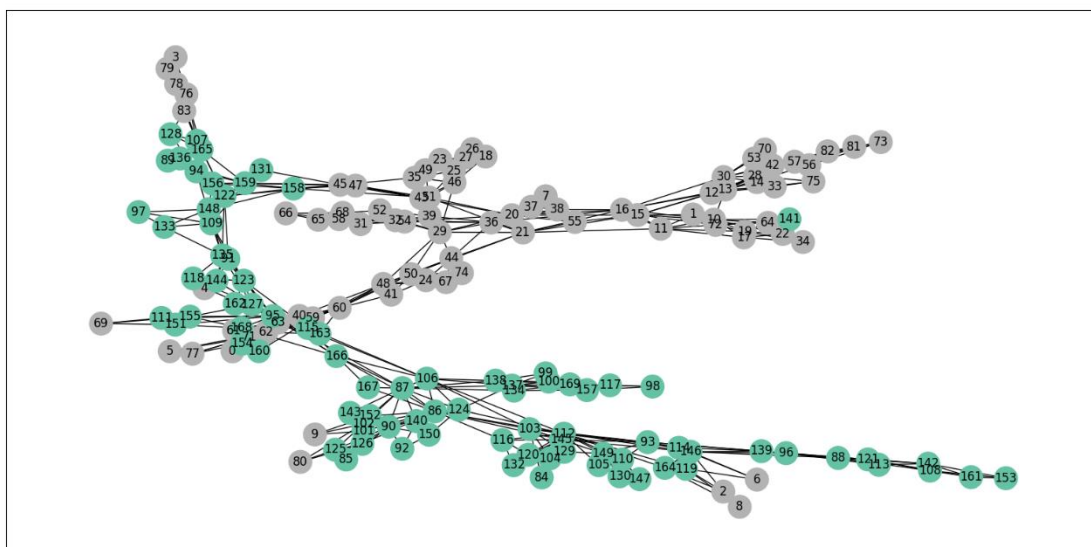
בשביל יצירת הרשתות עבור החלק השני של הפרויקט, שבהן הצמתים הן הזוגות, נוריד את השאלות היתירות ונשאיר רק שאלות מרכזיות אשר באמצעותן נוכל לבנות קשרים קורלטיביים בין התשובות של הזוגות על שאלות אלה ולזהות מאפיינים פסיכולוגים ממספר שאלות מצומצם ומזוקק יותר. את השאלות המרכזיות והמשמעותיות חיפשנו ברשתות של חלק א' באמצעות מיצוע ערכי  $\text{degree centrality}$  ו  $\text{eigenvector centrality}$  (לאחר נרמול לפי שיטת minmax) על כל אחד מהצמתים בכל קהילה ובחירת ערכי המינימום על גבי הקהילה.

שאפנו לקחת את הצמתים עם ערכי  $\text{eigenvector centrality}$  הנמוכים ביותר מפני שצומת שמחוברת לצמתים בעלי ערכי  $\text{centrality}$  נמוך יותר סביר שיהיה בהם מידע בעל שונות גבוה יותר מאשר קודקודים המקושרים לקודקודים מרכזיים, בחרנו בשיטה זו כי המטרה שלנו היא להגדיל את השונות. בנוסף ניתן לשים לב כי בהסתכלות על הקהילות עצמן, הערכים של ה  $\text{betweenness centrality}$  יחסית דומים אחד לשני והוספת המדד לא סייעה לשימור השונות הגבוהה ולכן ויתרנו עליו.

השאלות המשמעותיות לפי המדד מופיעים בנספח (1)

## חלק ב:

הסתכלנו על הרשת שבה האנשים הם הצמתים והקשתות של השאלות נקבעות רק לפי 15 השאלות שמצאנו. הצבע של הקודקוד נקבע על ידי האם הזוג הוא נשוי או גרוש, כאשר צומת אפורה מייצגת זוג גרוש וצומת ירוקה מייצגת זוג נשוי. מכיוון שהתשובות של הזוגות על השאלות המשמעותיות שמצאנו בחלק א' מהוות פרוקסי למאפיינים פסיכולוגיים של הזוגות, אם נוכל ליצור הפרדה "טובה" ומשמעותית בין זוגות נשואים לזוגות גרושים על הרשת שלנו, נוכל להסיק כי יש הבדלים פסיכולוגיים בין זוגות גרושים לזוגות נשואים. בסרטוט ניתן לראות כי ישנה הפרדה בלשהי בין הקבוצות אך הפרדה זו אינה משולמת.



כדי למדוד את טיב חלוקת הרשת לקהילות על סמך הלייבל, חישבנו את המודולריות על החלוקה שביצענו וקיבלנו 0.43. אמנם ערך זה גדול מ-0, הוא לא קרוב במיוחד ל-1 ולכן זה לא חד משמעי שההפרדה על סמך הלייבל היא "טובה".

כדי שנוכל לבחון את קיומם של הבדלים פסיכולוגיים בין זוגות גרושים לזוגות נשואים עשינו את ארבעת מבחני ההשערות הבאים:

1. האם ערך ה-ACC של זוגות גרושים שונה מערך ה-ACC של זוגות נשואים? השערת האפס שלנו היא שאין הבדל בעוד שההשערה האלטרנטיבית טוענת אחרת.
2. האם ערך ה-AEC של זוגות גרושים שונה מערך ה-AEC של זוגות נשואים? השערת האפס שלנו היא שאין הבדל בעוד שההשערה האלטרנטיבית טוענת אחרת.
3. האם ערך ה-ASPL של זוגות גרושים שונה מערך ה-ASPL של זוגות נשואים? השערת האפס שלנו היא שאין הבדל בעוד שההשערה האלטרנטיבית טוענת אחרת.



4. האם ערך המודולריות שקיבלנו על ידי חלוקת הרשת לקהילות גרושים ולקהילת נשואים הוא המקסימלי ביחס לערכי המודולריות של חלוקות אחרות של הרשת לשתי קבוצות צמתים? השערת האפס שלנו שלא בעוד שההשערה האלטרנטיבית טוענת אחרת.

עבור השוואות 1-3 חישבנו עבור כל צומת את הערך של המדד על הצומת, כדי שנוכל לקבל שתי דגימות להשוואה. מכיוון שמספר הזוגות הנשואים ומספר הזוגות הגרושים לא מספיק גדול בשביל להשתמש במבחנים המניחים שההתפלגות של ערכי המדדים בהשוואות 1-3 עבור כל קבוצה מתכנסים להתפלגות נורמלית לפי חוק המספרים הגדולים, השתמשנו במבחן הא-פרמטרי Kruskal-Wallis H. עבור השוואה 4, יצרנו 300 חלוקות שונות של הרשת לשתי קבוצות צמתים לפי אלגוריתם Kernighan–Lin, וחישבנו עבור כל חלוקה את המודולריות שלה. מכיוון שגודל הדגימה גדול יותר, השתמשנו במבחן t עם רמת מובהקות 0.05.

	ACC	AEC	ASPL
Kruskal-Wallis H Test Pvalue	0.450278	0.854098	2.164301e-14

ההשערות האפס של השוואות 1-2 לא נדחו. מכאן ניתן להסיק כי על כך שזוגות בשתי הרשתות נוטים ליצור קבוצות הדוקות שבהן יש רמה גבוהה של דמיון או תאימות בין הזוגות על סמך תשובותיהם על השאלות המשמעותיות. זה מצביע על כך שהזוגות בתוך הרשתות חולקים מאפיינים משותפים או בעלי דינמיקה דומה של יחסים, אשר באות לידי ביטוי במענה דומה על השאלות. ממצא זה שולל את האפשרות כי קיימים הבדלים פסיכולוגים בין זוגות גרושים לבין זוגות נשואים.

לעומת זאת השערת האפס של השוואה 3 נדחתה, כלומר ממוצע מרחק המסלול הקצר ביותר אצל הזוגות הגרושים שונה מממוצע מרחק המסלול הקצר ביותר אצל זוגות נשואים. בפרט, הערך שנצפה אצל הזוגות הגרושים גדול משמעותית מזה שנצפה אצל הזוגות הנשואים. ניתן להסיק מכך שמספר הזוגות הגרושים שענו בצורה דומה על השאלות המשמעותיות שאספנו היה גדול יותר בהשוואה אצל הזוגות הנשואים, מה שיכול להצביע על מעט הבדלים פסיכולוגיים בין הזוגות בתוך קהילת הזוגות הגרושים. אולם, ממצא זה אינו מספיק כדי להוכיח כי יש הבדלים פסיכולוגים בין זוגות נשואים לבין זוגות גרושים מכיוון שהוא רק מצביע על הימצאות ניכרת של הבדלים פסיכולוגים בין זוגות נשואים לבין עצמם לעומת זוגות גרושים לבין עצמם.

יתר על כן, השערת האפס של השוואה 4 לא נדחתה ( $pvalue = 1$ ), כלומר החלוקה של הצמתים ברשת לשתי קהילות על סמך הלייבל אינו מניב את ערך המודולריות המיטבי ביחס לשאר החלוקות האפשריות לשתי קהילות של צמתים על סמך אלגוריתם Kernighan – Lin. מכאן ניתן להסיק כי

אף על פי שערך המודולריות שהתקבל אינו נמוך ובפרט שלילי, עצם קיומם של הבדלים מובהקים בתשובות של הזוגות הנשואים והזוגות הגרושים על השאלות תלוי בספק רב.

## מסקנת המחקר

בעקבות תוצאות המבחנים הסטטיסטיים שבצענו בחלק השני, לא נמצאה עדות מובהקת סטטיסטית להבדלים פסיכולוגית בין זוגות נשואים לבין זוגות גרושים. תוצאה זו מנוגדת להשערת המחקר.

כחלק ממטרות המחקר, התעניינו בגורמים להצלחה ולכישלון הזוגי. ניכר מהשאלות המשמעותיות שמצאנו, כי המכנה המשותף הגדול ביותר שלהם הוא בעיקר וכמעט באופן בלעדי תחום התקשורת הבין אישית ולכן באופן לא מפתיע, אנו מבינים כי זהו גורם משמעותי בטיב הצלחת או כשלון הקשר.

## צעדים נוספים ומחקר עתידי

פרויקט זה היה מחקר מקדים בנושא של חקר קשרי נישואין אשר החל בבדיקה של אבני היסוד של נישואין בפועל, ע"י בדיקת התנהגות ואמונות אשר מתקיימות בפועל אצל הזוגות ובדיקת ההשלכות של פעולות אלה על הסטטוס הזוגי של הזוגות. במחקר עתידי שימשיך מחקר זה היינו רוצים לבדוק מדוע אבני יסוד אלו דווקא הם החשובים לאדם ולהצלחת הנישואים, ובנוסף להבין וללמוד מה הם דפוסי ההתנהגות והחשיבה אשר הובילו לסיום של קשרים זוגיים וגירושים. כלומר, נרצה לקחת את השאלות שנמצאו כרלוונטיות למידול התופעה וליצור שאלון יותר ממוקד על הנושא תוך התייעצות עם מומחה בתחום הפסיכולוגיה אשר יוכל לתת דגשים ממוקדים למידול התופעה. לאחר יצירת השאלון החדש נוכל לבצע את הפעולה שוב ולמצוא מיקוד אפילו יותר מתומצת של המאפיינים הנדרשים בתוך קשר על מנת שיצליח\ לא יצליח. בנוסף, ניתן לבחון איך התאמה בין דעות בין שני אינדיבידואלים – בני הזוג, העונים על שאלות אלו, משפיעה על טיב הקשר שלהם – האם אולי העניין הוא חוסר תיאום בנושא באותו (כמו שנהוג לומר "צריך שניים לטנגו") שמע אולי ישנה הטיה מסוימת בכך שהנתונים נאספו כאשר הזוג יושב ביחד וענה "ביחד"? אולי אחד מבני הזוג היה דומיננטי יותר על השני? בשביל לבדוק זאת נשתמש בשאלות של המחקר הקודם, רק שהפעם נמדל את השאלות כך שתהיה השאלה "האם אני מסכים עם זוגתי\זוגתי לשעבר?" על כל אחת מהן בשאלון המקורי. לאחר איסוף נתונים אלה, נשתמש בקורלציית פירסון (על מנת לדעת אם האם הייתה הסכמה חיובית או שלילית או חוסר הסכמה) על מנת לבדוק את הדמיון בשאלות בין הזוגות על מנת לייצר מכל זוג וקטור תשובות אחד (כלומר נחזור לתצוגה של וקטור לזוג) אשר ייצג את רמת ההסכמה על כל שאלה ולא את התוכן המענה על השאלה. ניקח את

הקשתות בעלות הקורלציות הכי משמעותיות (שליליות ביותר וחיוביות ביותר תוך שמירה על הסימן) ומכאן להמשיך הניתוח דומה (אך שונה מכיוון שהרשת אינה בינארית) מבחינת השלבים והמטרות להשגה לניתוח שנעשה במחקר הקודם בנוסף ניתן יהיה לשלב אלגוריתמים ייעודיים למציאת יתירות (אשר אנחנו לא הכרנו בתחילת העבודה על הפרויקט). ניתוח זה יוכל לחדד האם ההסכמה על השאלות הינה הגורם המשמעותי והאם בעצם בשאלות אלה מתחבאת משמעות נסתרת לכך שיש חוסר הסכמה\ראיה משותפת על נושאים אלה ולא דווקא בעיה עם הנושא עצמו, דבר אשר יכול לפתוח כיוון הסתכלות חדש לגמרי על הבעיה.

## ביבליוגרפיה

1. Yöntem, M , Adem, K , İlhan, T , Kılıçarslan, S. (2019). DIVORCE PREDICTION USING CORRELATION BASED FEATURE SELECTION AND ARTIFICIAL NEURAL NETWORKS. Nevşehir Hacı Bektaş Veli University SBE Dergisi, 9 (1), 259-273.  
(<https://dergipark.org.tr/tr/download/article-file/748448>)
2. Guido Previde Massara , T. Di Matteo Tomaso Aste (2015) [\*Network Filtering for Big Data: Triangulated Maximally Filtered Graph\*](#)
3. <https://www.kaggle.com/datasets/andrewmvd/divorce-prediction>

## נספחים:

(1) השאלות שנבחרו:

- רשת כל הזוגות:

שאלה 2: " I know we can ignore our differences, even if things get hard sometimes."

שאלה 3: " I know we can ignore our differences, even if things get hard sometimes."

שאלה 37: " My discussion with my spouse is not calm."

שאלה 43: " I mostly stay silent to calm the environment a little bit."

שאלה 52: " I wouldn't hesitate to tell my spouse about her/his inadequacy."

- רשת הזוגות הגרושים:

שאלה 6: " We don't have time at home as partners."

שאלה 11: " I think that one day in the future, when I look back, I see that my spouse and I have been in harmony with each other."

שאלה 12: " My spouse and I have similar values in terms of personal freedom".

שאלה 31: " I feel aggressive when I argue with my spouse"

שאלה 40: "We're just starting a discussion before I know what's going on."

שאלה 44: "Sometimes I think it's good for me to leave home for a while."

• רשת הזוגות הנשואים:

שאלה 7: "We are like two strangers who share the same environment at home rather than family."

שאלה 21: "I know exactly what my wife likes."

שאלה 48: "I feel right in our discussions."

שאלה 49: "I have nothing to do with what I've been accused of."