

מבחן ניתוח רשתות חברתיות

בחרתי לענות על 15 שאלות

שאלה 1

ראשית, טענתי את קבצי ה-csv ל-data frame, ביצעתי group by לפי מדינת המוצא והיעד וסיננתי מהם את הרשומות בהם שסכום המשקל היה מתחת ל-2 טון, או שהיו קיימים ערכים חסרים בעמודת ה-net weight. לאחר סינון זה, המרתי את ה-data frame לרשתות עבור כל אחת מהשנים.

סעיף א

על מנת למצוא את כמות המדינות שייבאו וייצאו חיטה בכל אחת מהשנים, ספרתי את מספר המדינות הייחודיות בעזרת פונקציות unique על עמודות ה-Reporter וה-Partner בכל אחת מהשנים. על מנת למצוא את כמות המדינות שגם ייצאו וגם ייבאו בכל אחת מהשנים השתמשתי בפונקציית intersect אשר מוצאת את החיתוך בין שתי רשימות המייבאים והמייצאים.

מספירה זו התקבלו עבור שנת 2016: 134 מדינות מייצאות, 223 מדינות מייבאות ו-133 מדינות שגם ייצאו וגם ייבאו בשנה זו.

עבור שנת 2021 התקבלו 124 מדינות מייצאות, 221 מדינות מייבאות ו-123 מדינות שגם ייצאו וגם ייבאו בשנה זו. בנוסף, התקבלו 110 מדינות שגם ייצאו וגם ייבאו חיטה, הן בשנת 2016 והן בשנת 2021.

ניתן לראות כי חלה ירידה הן בכמות המדינות המייצאות והן בכמות המדינות המייבאות בין השנים 2016 ל-2021.

```
[1] "2016:"
[1] "num of exporters in 2016: 134"
[1] "num of importers in 2016: 223"
[1] "num of countries that exported and imported in 2016: 133"
[1] "2021:"
[1] "num of exporters in 2021: 124"
[1] "num of importers in 2021: 221"
[1] "num of countries that exported and imported in 2021: 123"
[1] "num of countries that exported and imported in 2021 and in 2016: 110"
```

סעיף ב

על מנת למצוא את המדינות המרכזיות במערך הסחר בגרעיני חיטה ובמסחר במוצרי חיטה, ראשית חילקתי כל אחת מהשנים לשתי רשתות לפי סוגי המוצר המתאימים בהתאם לטבלה 1 במאמר. כך, התקבלו לכל שנה 2 טבלאות, האחת לגרעיני חיטה והשנייה למוצרי חיטה. את הטבלאות המרתי לרשתות ועליהן בחנתי את מדדי המרכזיות של המדינות. כחלק משלבי העיבוד המקדים, ביצעתי group by לפי מדינת המוצא, מדינת היעד וקוד המוצר, וזאת על מנת לקבל רשומה נפרדת עבור כל סוג מוצר, כאשר היא סכימה של כל פעולות הייצוא עבורו בין שתי המדינות.

בפלט הבא ניתן לראות את חמש המדינות המובילות מבחינת betweenness, degree, closeness ברשתות גרעיני החיטה ומוצרי החיטה של כל אחת מהשנים.

המשמעות של מדינות מרכזיות מבחינת closeness היא מדינות אשר מסלולי הייבוא מהן לשאר המדינות הם הקצרים ביותר, מדינות מרכזיות מבחינת degree הן מדינות אשר מייצאות הכי הרבה למדינות אחרות, ומדינות מרכזיות מבחינת betweenness הן מדינות אשר נמצאות על הכי הרבה מסלולי ייבוא/ייצוא קצרים ביותר בין שתי מדינות. ניתן לראות כי ארה"ב מובילה ב-betweenness בארבע הרשתות, וצרפת מובילה ב-degree בכולן.

2021 גרעיני חיטה	2021 מוצרי חיטה	
ארה"ב, רוסיה, צרפת, גרמניה, קנדה	ארה"ב, טורקיה, צרפת, איטליה, איחוד האמירויות	Betweenness
רוסיה, צרפת, EU2, גרמניה, ארה"ב	איטליה, EU2, צרפת, גרמניה, טורקיה	Degree
הרפובליקה הדומיניקנית, הונדורס, ג'מייקה, קירגיסטן, ערב הסעודית	ברבדוס, גרנדה, איסלנד, קירגיסטן, ירדן	closeness

2016 גרעיני חיטה	2016 מוצרי חיטה	
ארה"ב, גרמניה, צרפת, רוסיה, קנדה	ארה"ב, טורקיה, צרפת, איחוד האמירויות, דרום אפריקה	Betweenness
גרמניה, צרפת, רוסיה, EU2, ארה"ב	צרפת, גרמניה, טורקיה, EU2, איטליה	Degree
בליז, גאנה, מרוקו, ערב הסעודית, טרינאדד	גינאה, איסלנד, סמואה, טג'יקיסטן, EU2	closeness

```
[1] "----2021 cereal:---"
[1] "top 5 countries by betweenness:"
      USA      RUS      FRA      DEU      CAN
1937.6239 1530.5865 1048.0937 890.6459 782.4609
[1] ""
[1] "top 5 countries by degree:"
      RUS      FRA      EU2      DEU      USA
101 93 88 84 83
[1] ""
[1] "top 5 countries by closeness:"
      DOM      HND      JAM      KGZ      SAU
1 1 1 1 1
[1] ""
[1] "----2021 wheat products:---"
[1] "top 5 countries by betweenness:"
      USA      TUR      FRA      ITA      ARE
5538.065 2890.573 2239.349 1930.819 1825.534
[1] ""
[1] "top 5 countries by degree:"
      ITA      EU2      FRA      DEU      TUR
156 153 153 150 149
[1] ""
[1] "top 5 countries by closeness:"
      BRB      GRD      ISL      KGZ      JOR
1.0 1.0 1.0 1.0 0.5
```

```
[1] "----2016 cereal:---"
[1] "top 5 countries by betweenness:"
      USA      DEU      FRA      RUS      CAN
2433.409 1340.578 1278.260 1036.619 1032.589
[1] ""
[1] "top 5 countries by degree:"
      DEU      FRA      RUS      EU2      USA
108 103 101 93 88
[1] ""
[1] "top 5 countries by closeness:"
      BLZ      GHA      MAR      SAU      TTO
1 1 1 1 1
[1] ""
[1] "----2016 wheat products:---"
[1] "top 5 countries by betweenness:"
      USA      TUR      FRA      ARE      ZAF
5825.047 3511.115 2661.693 2076.360 1895.824
[1] ""
[1] "top 5 countries by degree:"
      FRA      DEU      TUR      EU2      ITA
155 151 151 147 138
[1] ""
[1] "top 5 countries by closeness:"
      GIN      ISL      WSM      TJK      EU2
1.000000000 1.000000000 1.000000000 0.500000000 0.003246753
```

סעיף ז

על מנת לבדוק האם סביר שקשר משולשי יהיה פחות נפוץ ברשת ייצוא/ייבוא, יש להסתכל על מפקד השלשות של השלשה המתאימה לקשר משולשי כזה, אשר מספרה במפקד השלשות הוא 10, ולראות האם כמות השלשות מסוג זה ברשת שלנו דומה לממוצע כמות השלשות האלה ברשתות רנדומליות אחרות.

לצורך כך, הגרלתי 100 רשתות רנדומליות וחישבתי עבורן את מפקד השלשות. לאחר מכן, ביצעתי מבחן t test עבור כמות השלשות המעגליות ברשת שלנו למול ממוצע כמות השלשות המעגליות ב-100 הרשתות שדגמנו, כאשר ההשערה היא שברשתות הרנדומליות ממוצע כמות השלשות הללו יהיה גדול מאשר ברשת הייצוא/ייבוא. את המבחן ביצעתי הן על רשת הייצוא/ייבוא של שנת 2021 והן של הרשת של שנת 2016.

[1] "2021:"

One Sample t-test

```
data: tcRand2021[, 10]
t = 233.77, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 81
95 percent confidence interval:
 809.8064      Inf
sample estimates:
mean of x
 815.02
```

[1] "2016:"

One Sample t-test

```
data: tcRand2016[, 10]
t = 234.78, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 80
95 percent confidence interval:
 774.5133      Inf
sample estimates:
mean of x
 779.46
```

בשני המבחנים הללו קיבלנו ערך pvalue קטן מ-0.05 אשר מציין כי

עבור שתי השנים, הקשר המשולשי ברשתות הייצוא/ייבוא פחות

סביר ונפוץ בהשוואה לרשתות רנדומליות אחרות וההשערה אכן

נתמכת.

שאלה 2

ראשית, טענתי את קובץ הנתונים וסיננתי ממנו קשתות בעלות משקל של 0. הסיבה לסינון זה היא שמשקל הקשתות מציין את מספר שיתופי הפעולה בין האזורים, ועל כן קשתות שמשקלן 0 משמעותן שלא קיים שיתוף פעולה, ועל כן הקשת מיותרת. לאחר הסינון, הפכתי את ה-data frame לרשת.

סעיף א

על מנת לבדוק האם הרשת נטולת סקאלה, הרצתי מבחן `fit_power_law`. ממבחן זה התקבל כי הרשת נטולת סקאלה החל מדרגה 15. עבור ערך זה, ה-pvalue גדול ומתקרב ל-1 (0.93) ועל כן נוכל לומר שהרשת נטולת סקאלה החל מדרגה זו, וערך האלפא המתאים הוא 3.23. אם נבדוק את הרשת החל מדרגה 1 נקבל שהרשת החל מדרגה זו אינה נטולת סקאלה מאחר וערך ה-pvalue המתקבל קטן מ-0.05.

קיבלנו שהרשת נטולת סקאלה החל מדרגה 15, המשמעות היא שהחל מדרגה זו ישנם אזורים בעלי דרגה גבוהה מהיתר באופן משמעותי, כלומר מאוד מרכזיים ברשת שיתופי הפעולה ביחס לשאר. המשמעות של רשת נטולת סקאלה בהקשר לרשת שיתופי הפעולה, הוא שקיימים מספר אזורים מועט אשר נמצאים בשיתוף פעולה עם כמות גדולה יחסית של אזורים אחרים, ומרבית האזורים משתפים פעולה עם מספר קטן יחסית לזה של אזורים אחרים.

```
$continuous
[1] FALSE

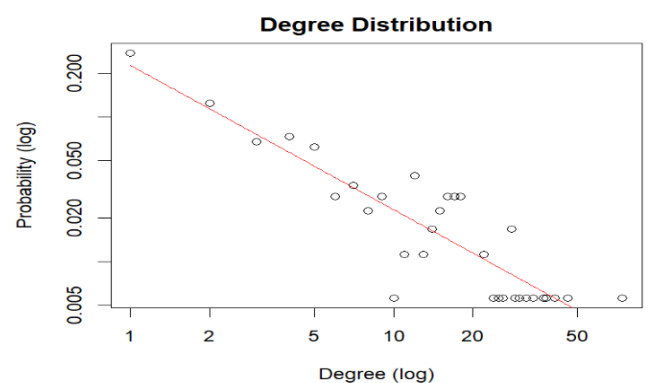
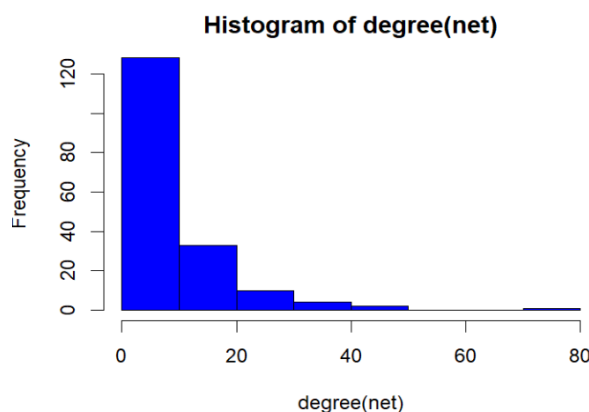
$alpha
[1] 3.233689

$xmin
[1] 15

$logLik
[1] -129.4896

$KS.stat
[1] 0.08642185

$KS.p
[1] 0.9327785
```



סעיף ב

על מנת לבחון האם הרשת היא רשת עולם קטן, יש להסתכל על יחס מקדם ההתקבצות ועל יחס ממוצע הגאודזים בין הרשת שלנו לרשת רנדומלית. על יחס מקדם ההתקבצות להיות גדול מ-10, ויחס ממוצע הגאודזים להיות בערך 1.

מבדיקה זו קיבלנו יחס מקדם התקבצות של 4.920608, כלומר קטן מ-10, ויחס ממוצע גאודזים 1.02415 הקרוב מאוד ל-1. מאחר ואנו לא עונים על התנאי הראשון, נאמר כי רשת שיתופי הפעולה אינה רשת עולם קטן, כלומר שהאזורים השונים ברשת שיתופי הפעולה אינם מחוברים חזק האחד לשני במונחים של שיתופי פעולה והמשמעות

היא שעל מנת לקשר בין כל שני אזורים שונים בעזרת שיתופי הפעולה הקיימים ברשת, יש לעבור דרך מספר רב יחסית של אזורים אחרים.

```
[1] 4.920608
[1] 1.02415
```

סעיף ג

על מנת לבחון האם רשת שיתופי הפעולה היא רשת ארדוש-רייני, נגריל רשת ארדוש-רייני בעלת אותו מספר צמתים וקשתות כמו ברשת שיתופי הפעולה, ונבחן בעזרת מבחן קולמגורוב-סמירנוף האם הרשתות מגיעות מאותה התפלגות דרגה. ממבחן זה קיבלנו ערך p -value קטן מ-0.05 ועל כן נדחה את השערת האפס שטוענת שהרשתות מגיעות מאותה התפלגות דרגה ונאמר שהרשת שלנו היא לא רשת ארדוש-רייני.

קיבלנו שהרשת היא לא ארדוש רייני, כלומר שהקשרים בין האזורים השונים אינם אקראיים. המשמעות של כך היא שייתכן וקיימים גורמים שונים או דפוסי התחברות הגורמים לשני אזורים להתחבר ביניהם בעזרת שיתוף פעולה והדבר אינו אקראי. גורמים אלה יכולים להיות דמיון במאפייני האזור, קרבה גאוגרפית או דמיון בדרך ההתנהלות.

Two-sample Kolmogorov-Smirnov test

data: igraph::degree(erdos) and degree(net)
D = 0.44944, p-value = 4.441e-16
alternative hypothesis: two-sided

סעיף ד

ממצאי הסעיפים הקודמים, קיבלנו שהרשת היא לא רשת נטולת סקאלה, לא רשת עולם קטן ולא רשת ארדוש רייני. בהקשר לזרימת המידע בין חברות הביטק ניתן לומר כי מאחר והרשת היא לא רשת עולם קטן, זרימת המידע תהיה איטית ברשת, מאחר ועל מנת שמידע יעבור בין שני אזורים שונים יהיה עליו לעבור דרך מספר יחסית רב של אזורים אחרים כי אורך המסלול הקצר ביותר הממוצע בין קודקודים הוא יחסית ארוך.

הרשת הא לא רשת נטולת סקאלה, כלומר לרשת יש התפלגות הומוגנית יותר של דרגות הצמתים בהשוואה לרשתות נטולות סקאלה. מאחר והרשת לא נטולת סקאלה, פיזור המידע עשוי להיות איטי יותר מאחר וקיימים פחות אזורים המובילים בהשפעתם ביחס לאחרים בהקשר להפצת המידע.

מאחר והרשת היא לא רשת ארדוש רייני, כלומר התפלגות הדרגה אינה אקראית, ניתן לשער כי קיימים גורמים אחרים אשר משפיעים על יצירת שיתוף פעולה בין האזורים, דבר אשר עשוי לעכב את זרימת המידע ברשת. ניתן להניח כי גורמים אלו מובילים ליצירת קשר של שיתוף פעולה בין אזורים דומים או בעלי מאפיינים משותפים, דבר אשר מגביל את יכולת ההפצה של המידע ברשת לאזורים אחרים, ועלול להשאיר את המידע בסביבתו הקרובה של הארגון ולהקטין את מהירות העברת המידע.

סעיף ה

בהתבסס על רכיב הקשירות הגדול של הגרף, אלגוריתמי החלוקה השונים מביאים חלוקה אחרת לקהילות. על מנת לבחור את החלוקה המיטבית לקהילות על רכיב זה, הרצתי אלגוריתמי חלוקה לקהילות של Louvain, edge betweenness, fast greeasy, cluster leasing eigen, והשווייתי ביניהם בעזרת ערך המודולריות המתקבל מהחלוקה.

החלוקה אשר הביאה לערך המודולריות הגדול ביותר היא חלוקה לפי Louvain, אשר הניבה 9 קהילות וערך מודולריות של 0.33. לכן, כדאי לחלק את הרשת ל-9 קהילות, לפי החלוקה לקהילות שקיבלנו באלגוריתם זה.

```
> #louvain
> cl<-cluster_louvain(largestComp, resolution=1)
> cl
IGRAPH clustering multi level, groups: 9, mod: 0.33
+ groups:
$`1`
[1] "1003" "1051" "1057" "8116" "9182" "9761" "11000" "12061" "13001" "13004" "14292" "1002" "1001"
[14] "7138" "5113" "8115" "9180" "9679"

$`2`
[1] "2000" "3152" "4012" "5766" "5974" "6414" "6434" "6531" "9174" "9188" "9362" "9775" "9779"
[14] "10045" "12065" "12069" "13051" "13062" "1058" "3456" "3454" "3460"

$`3`
[1] "3155" "6534" "8118" "8212" "8221" "8222" "8236" "8311" "8315" "8317" "8326" "8415" "8417"
+ ... omitted several groups/vertices
> modularity(cl)
[1] 0.3301233
```

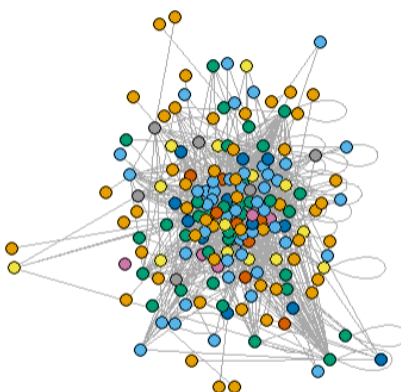
סעיף ו

k-core הוא תת גרף בו כל צומת שכן ללפחות k צמתים בתת הגרף. בחלוקת הגרף לקהילות בעזרת התקבלו 9 קהילות, בעוד שבחלוקה לקהילות בעזרת k-core התקבלו 11 קהילות. יחד עם זאת, אלגוריתם Louvain מתחיל בכל פעם בצורה אקראית אחרת ועשוי להוביל לחלוקה למספר אחר של קהילות בכל הרצה, כך וייתכן ובהרצות אחרות נקבל גם בעזרת Louvain 11 קהילות. שני האלגוריתמים מבצעים חלוקה לקהילות בצורה שונה. בעוד ש-Louvain פועל בצורה חמדנית בכל איטרציה למיקסום המודולריות, כלומר למקסם את כמות הקשרים בתוך כל קהילה ביחס לכמות הקשרים מחוצה לה, אלגוריתם k-core כלל לא מתייחס למודולריות ולא ניתן לחשב עבור החלוקה המתקבלת ממנו ערך זה. ניתן לראות את ההבדלים בחלוקה לקהילות בגרף מטה.

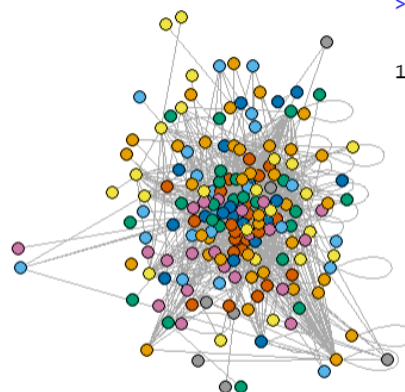
מבחינת גודל הקהילות, ניתן לראות כי בהתאם למצופה, גדלי הקהילות המתקבלות שונים בכל אחד מהאלגוריתמים. נראה כי החלוקה שונה מהותית, והיא לא פירוק של קהילה לתתי קהילות קטנות יותר.

על כן, ניתן לומר שהחלוקה לקהילות שונה הן מבחינת דרך החלוקה, הן במספר הקהילות המתקבלות והן בגודלן.

K-core



Louvain

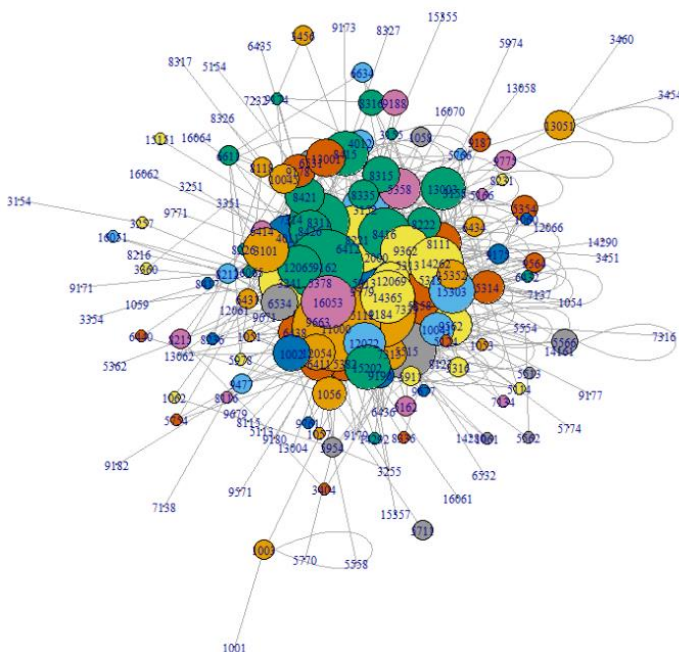


```
> kcores_sizes
kcores
 1  2  3  4  5  6  7  8  9 10 11
48 23 15 14 12  8  4  7  2 23 20
> louvain_sizes
 1  2  3  4  5  6  7  8  9
17 27 31 29 17  8 20 12 15
```

סעיף ז

ההחזיה שבחרתי לרשת מבוססת על תכונת ה-act, כלומר על מספר חברות ומוסדות הביוטכנולוגיה באזור. הצבעים של הצמתים השונים בגרף הם לפי החלוקה לקהילות המתקבלת בעזרת אלגוריתם Louvain, וגודל הצמתים הוא הטרנספורמציה הלוגריתמית של מספר חברות הביוטכנולוגיה באזור (נעשתה טרנספורמציה זו על מנת שהגרף יהיה קריא וברור מאחר וערכי שדה זו שונים משמעותית בין הצמתים השונים). בחזרתי להתייחס דווקא למאפיין זה מאחר וכמות חברות הביוטכנולוגיה באזור מסוים עשויות להצביע על "החוזק שלו" כאזור משמעותי, וייתכן וגם על כמות שיתופי הפעולה שייצור מאחר וככל שיש יותר חברות באזור מסוים, סביר כי כמות שיתופי הפעולה תהיה גדולה יותר. באמצעות סימון תכונה זו כגודל של צומת בגרף, ניתן לזהות את הצמתים המרכזיים מבחינה זו.

מהגרף המתקבל ניתן לראות כי השערה זו יחסית מתקבלת, וכי ניתן לזהות כי אזורים בעלי שיתופי פעולה רבים (כמות קשתות גדולה) מופיעים בגרף גם כצמתים בעלי רדיוס גדול (כלומר אזורים בעלי הרבה חברות ביוטק), בעוד שצמתים בעלי מספר שיתופי פעולה קטן מופיעים בגרף כצמתים עם רדיוס קטן.



שאלה 3

סעיף א

המשתתפים ב-SUS וב-RC הוזמנו למלא סקר אינטרנטי שמטרתו להעריך את קיומם ואופיים של הקשרים ברשתות המעורבות מתוך הנחה שהם קשורים למניעת השמנת יתר בקרב ילדים.

המשתתפים בסקר התבקשו למלא מידע דמוגרפי ולאחר מכן לנקוב בשמות של עד 20 משתתפים אשר איתם דנו בסוגיות הקשורות להשמנת יתר בקרב ילדים (בקרב קבוצות המעורבות או מחוצה להן), ועבור כל קשר כזה לציין את מידת הקרבה ביחסים אל אותו אדם. מכיוון שהשאלון הועבר לאחר ההשתתפות בקבוצות המעורבות, ניתן למשתתפים מידע הקשרי אודות תקופת הזמן בה התקיימו הקבוצות, וכן הם התבקשו לחשוב על זמן ההתערבות ולתאר מאפיינים שונים אשר הם זוכרים מאותה התקופה וזאת על מנת לסייע לעורר את זכרונם.

לאחר המענה על הסקר, הקודקודים ברשת הוגדרו להיות משתתפי הסקר וכלל האנשים שצוינו על ידם, קשרי הדיונים שתיארו הוגדרו להיות הקשתות בגרף.

החיסרון בשיטה זו היא תהליך הדגימה הבעייתי. ראשית, לא כל המשתתפים בקבוצות המעורבות ענו על הסקר, דבר אשר יוצר דגימה חלקית בלבד מהרשת וייתכן ואם כלל המשתתפים היו מדווחים על הקשרים הללו, הייתה מתקבלת רשת מורכבת יותר ובעלת מידע רב יותר. דגימה זו מהווה הסתכלות חלקית על תת גרף מהרשת המלאה.

בנוסף, המשתתפים בסקר היו רק אלו אשר השתתפו בקבוצות המעורבות אך הם יכולו לדווח על גורמים אחרים מחוצה לה אשר היו עימם קשר בנושאים אלה והם צורפו לרשת. אותם גורמים לא מילאו בעצמם את הסקר, והדבר לא מתאר את הרשת במלואה ויוצר חוסר במידע אודות האנשים אשר הגורמים החיצוניים היו עימם בקשר. לגורמים החיצוניים לקבוצות המעורבות הללו יהיו קשתות נכנסות בלבד ללא קשתות יוצאות, דבר אשר לא בהכרח מתאר נכונה את המציאות.

כמו כן, השאלון מבוסס על זיכרונם של המשתתפים בנוגע לאנשים עימם היו בקשר. החיסרון של שיטה זו הוא שאחזור מהזיכרון לא בהכרח יהיה מדויק, וייתכן ועל אף השיטות השונות בהן נקטו על מנת לעורר את הזיכרון של המשתתפים, עדיין לא ידווחו כלל הקשרים אשר היו קיימים בפועל, דבר אשר עשוי להשפיע על מבנה הרשת המתקבל. יחד עם זאת, לשיטה יתרון במצבים בהם הפעולות הללו שננקטו דווקא יחדדו את הזיכרון ויגרמו למשתתפים להיזכר בקשרים אשר היו נאבדים אחרת.

למרות שיטת הדגימה אשר עלולה לא לייצג את המציאות בצורה נכונה, שיטה זו מאפשרת מידול של הרשת (גם אם חלקי) וייתכן שמבחינת מטרות המחקר הדבר מספק. מידול חלקי זה יכול להיחשב אף כיתרון מאחר והוא קל יותר לאיסוף בהשוואה למקרה בו היו דוגמים את הרשת במלואה, והוא עשוי לחסוך זמן, כסף ומשאבים.

סעיף ב

על מנת לנתח את הרשתות ברמת המיקרו, ננתח את מבנה הרשתות:

RC	SUS	
54	99	מספר צמתים
126	218	מספר קשתות
0.04402516	0.02246959	צפיפות
1	1	מספר רכיבי קשירות
3	3	קוטר
1.851415	2.142072	ממוצע גאודזים
0.06918239	0.0897753	In degree centralization
0.427673	0.4469182	out degree centralization
0.4054768	0.569471	Betweenness centralization
0.4653922	0.5123259	Closeness centralization

ניתן לראות כי על אף מספר הצמתים והקשתות השונה בשתי הרשתות, בשתייהן יש רכיב קשירות אחד, קוטר 3 ומדדי degree centralization דומים יחד עם זאת, ה-betweenness וה-closeness centralization מעט שונים בשתי הרשתות, כאשר הם מעט גדולים יותר ברשת ה-SUS.

מאחר וערך ה-betweenness centralization של רשת SUS גדול מעט יותר משל רשת ה-RC, ניתן לומר כי ברשת זו יש צמתים אשר נמצאים על יותר מסלולים קצרים בין שני צמתים אחרים ועשויים להוות "גשרים" או צמתים משמעותיים בהעברת המידע ברשת. בהתייחס לרשתות המתוארות, ייתכן ואלו אנשים מרכזיים אשר היוו גורמים מקשרים בשיחות בין אנשים שונים. גורמים אלו עשויים להיות מנחי הקבוצות, אנשי מפתח חיצוניים אשר עמדו בקשר עם המשתתפים או משתתפים מרכזיים אחרים, וניתן לומר כי ברשת ה-SUS היו גורמים מתווכים בין המשתתפים בצורה מעט משמעותית יותר. מאחר וערך ה-closeness centralization של רשת SUS גדול מעט יותר משל רשת ה-RC, ניתן לומר כי ברשת זו רמת הקרבה בין הצמתים לשאר הצמתים האחרים גדולה יותר. סביר להניח כי ברשת זו קיימת גישה רחבה יותר למידע מאחר והמרחק הגאודזי בין הצמתים קטן יותר.

בהקשר לקבוצות המעורבות, רשתות בעלות ערך קרבה גבוה הן רשתות אשר עשויות לספק תמיכה רחבה יותר למשתתפים, מאחר והן מאפשרות הגעה מהירה יותר לקבלת תמיכה ממשתתפי הקבוצה. שתי הרשתות בעלות רכיב קשירות יחיד, דבר אשר לא מותיר אף חבר בקבוצה לבדו, וכלל המשתתפים עשויים לקבל את התמיכה וההתערבות הנדרשת בתהליך.

סעיף ג

על מנת לבחון עד כמה התנהלות מיקרו זו יכולה לנבוע מאקראיות, נשווה את שתי הרשתות שלנו לרשתות רנדומליות שנגריל בעלות אותו מספר צמתים וקשתות כמו ברשתות המקוריות, ונשווה ביניהן בעזרת מבחן קולמגורוב-סמירנוף. מבחן זה מאפשר לבדוק האם הרשתות מגיעות מאותה התפלגות של רשת רנדומלית. מאחר ומקודם בחנו ארבעה מדדי מרכזיות שונים, נבחן האם הרשתות שלנו מגיעות מאותה התפלגות המדדים הללו כמו רשתות רנדומליות.

לצורך כך, עבור כל רשת הרצתי ארבע מבחני t: עבור התפלגות דרגה נכנסת, התפלגות דרגה יוצאת, betweenness ו-closeness. בכל שמונת המבחנים הללו ערכי ה-pvalue יצאו קטנים מ-0.05, ולכן נדחה את השערות האפס ונאמר ששתי הרשתות לא מגיעות מאותה התפלגות דרגה נכנסת, דרגה יוצאת, betweenness, ו-closeness כמו של רשת רנדומלית באותו הגודל, והתנהלותן ברמת המיקרו לא נובעת מאקראיות.

```
> ks_sus1
Two-sample Kolmogorov-Smirnov test

data: igraph::degree(randsus, mode = "in") and degree(sus, mode = "in")
D = 0.62626, p-value < 2.2e-16
alternative hypothesis: two-sided

> ks_sus2
Two-sample Kolmogorov-Smirnov test

data: igraph::degree(randsus, mode = "out") and degree(sus, mode = "out")
D = 0.82828, p-value < 2.2e-16
alternative hypothesis: two-sided

> ks_sus3
Two-sample Kolmogorov-Smirnov test

data: igraph::betweenness(randsus, directed = FALSE) and betweenness(sus, directed = FALSE)
D = 0.71717, p-value < 2.2e-16
alternative hypothesis: two-sided

> ks_sus4
Two-sample Kolmogorov-Smirnov test

data: igraph::closeness(randsus, mode = "all") and closeness(sus, mode = "all")
D = 0.52004, p-value = 8.184e-12
alternative hypothesis: two-sided

> ks_rc1
Two-sample Kolmogorov-Smirnov test

data: igraph::degree(randrc, mode = "in") and degree(rc, mode = "in")
D = 0.51347, p-value = 1.992e-08
alternative hypothesis: two-sided

> ks_rc2
Two-sample Kolmogorov-Smirnov test

data: igraph::degree(randrc, mode = "out") and degree(rc, mode = "out")
D = 0.84175, p-value < 2.2e-16
alternative hypothesis: two-sided

> ks_rc3
Two-sample Kolmogorov-Smirnov test

data: igraph::betweenness(randrc, directed = FALSE) and betweenness(rc, directed = FALSE)
D = 0.78114, p-value < 2.2e-16
alternative hypothesis: two-sided

> ks_rc4
Two-sample Kolmogorov-Smirnov test

data: igraph::closeness(randrc, mode = "all") and closeness(rc, mode = "all")
D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided
```

סעיף ד

על מנת לבחון באילו מהרשתות מידע יכול לעבור במהירות רבה יותר, נבחן את ממוצע הגאודזים של שתי הרשתות. ממוצע הגאודזים ברשת הוא ממוצע אורך המסלולים הקצרים ביותר בין כל שני קודקודים בגרף. מאחר ונכתב כי המידע עובר במסלולים קצרים חסרי כיווניות, נחשב את ממוצע הגאודזים כאשר מתעלמים מהכיווניות של הקשתות בגרף.

ניתן לראות כי ממוצע הגאודזים בשתי הרשתות תוך התעלמות מכיווניות הקשתות הוא די דומה, אך קצר יותר בעבור רשת ה-RC, ועל כן מידע יכול לעבור ברשת זו במהירות רבה יותר.

```
> average_path_lengthSUS <- average.path.length(sus, directed=FALSE)
> average_path_lengthSUS
[1] 2.570604
> average_path_lengthRC <- average.path.length(rc, directed=FALSE)
> average_path_lengthRC
[1] 2.349406
```

בנוסף, מסעיף ב ניתן לראות כי הצפיפות של רשת ה-RC גדולה פי 2 מזו של רשת ה-SUS, כלומר היחס בין מספר הקשתות בפועל לבין מספר הקשתות האפשריות גדול יותר ברשת ה-RC. בהקשר להעברת מידע, סביר כי ברשת עם צפיפות גבוהה המידע יתפשט מהר יותר מאחר ויש יותר נתיבים דרכם המידע יכול לעבור.

מדד נוסף אשר ניתן להסתכל עליו על מנת לבחון את מהירות העברת המידע ברשת הוא מדד ה-*eccentricity*. מדד ה-*eccentricity* של צומת הוא מידת החריגות שלו, כלומר מידת המרחק שלו מהצומת הרחוק ביותר ממנו. עבור כל אחת מהרשתות נחשב ערך מדד זה ונמצע אותו, על מנת לקבל את מדד החריגות הממוצע של הרשת. ניתן לראות כי ערך החריגות הממוצע גדול יותר ברשת ה-SUS, כלומר מידת החריגות של הצמתים בה גדולה יותר. פירוש הדבר הוא שהעברת מידע בין צומת מסוים לצומת הרחוק ביותר ממנו תדרוש יותר מעברים ברשת ה-SUS, דבר אשר מוביל לזרימת מידע איטית יותר ברשת.

```
> mean(eccentricity(sus, mode="all"))
[1] 3.686869
> mean(eccentricity(rc, mode="all"))
[1] 3.092593
```

מהסיבות הללו נאמר כי המידע יעבור במהירות רבה יותר ברשת ה-RC.

סעיף ה

על מנת להחליט אילו צמתים כדאי להוציא מכל רשת על מנת להגדיל את המסלולים הללו, יש להסתכל על הצמתים שמדד ה-*betweenness* שלהם הוא הגבוה ביותר בכל רשת. צמתים אלו נמצאים על מספר גדול של מסלולים קצרים ביותר בין קודקודים, ולכן אם נוציא צמתים אלו, המסלולים עלולים להתארך וזרימת המידע תהיה איטית יותר.

על מנת למצוא את הצמתים בעלי ה-*betweenness* הגדול ביותר בכל אחת מהרשתות, נחשב עבור כל צומת את ה-*betweenness* שלו ונמצא את מספרי הצמתים בעלי הערכים הגדולים ביותר במדד זה.

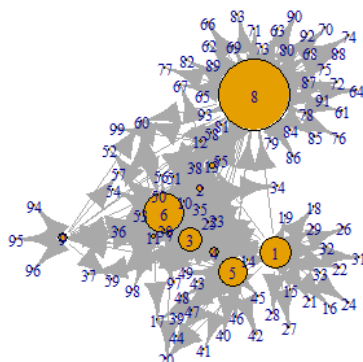
```
> top_betweenness_sus
[1] 1 3 5 6 8
```

ניתן לראות כי עבור רשת ה-SUS צמתים אלו יהיו 1,3,5,6,8, ועבור רשת ה-RC צמתים אלו יהיו 1,2,3,5,8.

```
> top_betweenness_rc
[1] 1 2 3 5 8
```

ניתן לראות גם את ה-*betweenness* של הצמתים בשתי הרשתות בגרף הבא. עבור כל צומת, גודלו הוגדר להיות בהתאם לערך ה-*betweenness* שלו. אכן ניתן לראות כי הצמתים שצוינו לעיל הם הצמתים בעלי ערכי ה-*betweenness* הגדולים ביותר בכל אחת מהרשתות, והם מרכזיים מבחינת מדד זה בפער גדול ביחס לשאר הצמתים בגרף.

SUS Betweenness



RC Betweenness

