**מטלה 1**

**תיאור הקוד:**

יצרנו שלושה מבני נתונים ומחלקה על מנת לממש את ה-API:

Graph – מילון אשר כל key הוא id של קודקוד בגרף ו-value הוא אובייקט Node של אותו קודקוד.

Mapper – מילון אשר ממפה בין name (ה-key במילון) של הקודקוד לבין ה-id (ה-value במילון) שלו.

page\_rank\_list – רשימה אשר מכילה tuples של שם הקודקוד וה-pageRank שלו בצורה הבאה: (node\_name, pageRank).

Node – מחלקה המתארת קודקוד ומכילה את השדות הבאים:

Name – שם הקודקוד

list\_of\_in – רשימה המכילה ids של כל הקודקודים שמצביעים לקודקוד

list\_of\_out – רשימה המכילה ids של כל הקודקודים שהקודקוד מצביע אליהם

page\_rank – ערך ה-pageRank של הקודקוד

prev\_pr – ערך ה-pageRank באיטרציה הקודמת.

בפונקציה load\_graph עבור כל קשת הרשומה בקובץ ה-csv הנתון נבדוק האם קודקודי ה-source וה-destination קיימים במבני הנתונים – אם לא – ניצור את הקודקודים ונוסיפם למבני הנתונים ולאחר מכן נוסיף את קודקוד המקור לרשימת list\_of\_in של קודקוד היעד ונוסיף את קודקוד היעד לרשימת list\_of\_out של קודקוד המקור.

בפונקציה calculate\_page\_rank, על פי האלגוריתם, תחילה מאתחלים את הpageRank של כל הקודקודים ב-1/N כאשר N הוא מספר הקודקודים בגרף. לאחר מכן ישנו for ראשי הרץ על כל הקודקודים ב-graph ועבור כל קודקוד מעדכנים את prev\_pr להיות page\_rank ורצים בfor על כל ה-list\_of\_in – כלומר על כל הקודקודים שמצביעים עליו. על פי ה-prev\_pr שלהם, גודל הרשימה list\_of\_out שלהם וה-b הנתון מחשבים את ה-page\_rank עבור הקודקוד ומעדכנים אותו באובייקט ב-graph. רצים על כל הקודקודים עד אשר תנאי העצירה מתקיים – כאשר עברו 20 איטרציות או כאשר סכום ההפרשים page\_rank -prev\_pr של כל הקודקודים קטן מ-e נתון. בסיום הפונקציה ממיינים את ה-graph אל תוך הרשימה page\_rank\_list מהpage\_rank הגבוה לנמוך.

בפונקציה get\_PageRank באמצעות ה-mapper נקבל את ה-id של האובייקט ונחזיר את הpage\_rank שלו מה-graph באופן מיידי.

בפונקציה Get\_top\_nodes נחזיר רשימה עם n הקודקודים הראשונים מתוך הרשימה page\_rank\_list.

בפונקציה get\_all\_PageRank נחזיר את כל הרשימה page\_rank\_list שכאמור ממויינת כבר.

**הרשתות:**

1. 2meirl4meirl הוא sub-reddit מהאתר reddit. קיבלנו קובץ המתאר רשת מתוך אתר זה ממסטרנט במחלקה. כל קודקוד ברשת הוא יוזר וכל קשת קיימת כאשר יוזר אחד הגיב ליוזר אחר, כלומר תהיה קשת מהיוזר שהגיב ליוזר שהגיב לו.

הרשת מורכבת מ10675 קודקודים ומ28288 קשתות.

עשרת הקודקודים עם ה-pageRank הגבוה ביותר הם:

('DuhPai', 0.00702889176400032)

('JackD2000', 0.004974128636944055)

('Arorise', 0.00488143120871658)

('snakeoil-huckster', 0.004566030155767194)

('Tyler1492', 0.004280961302821606)

('DelusionalHope', 0.00408104632538783)

('TheVeryBakedPotato', 0.003911526421686408)

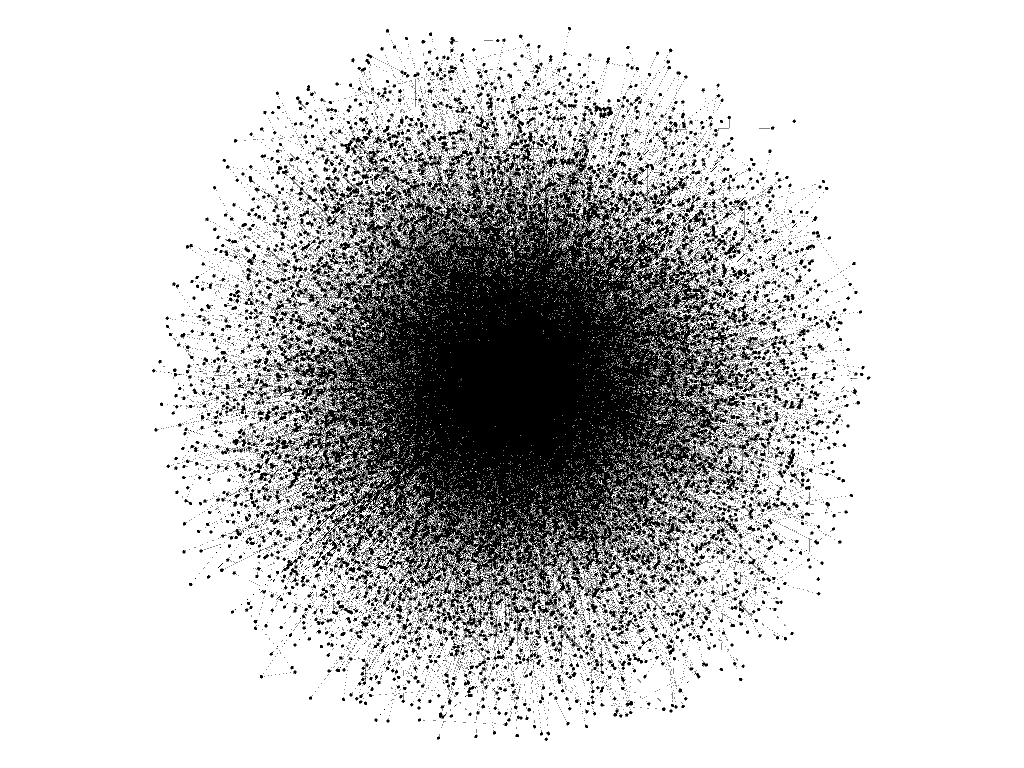
('jallopsmith', 0.003200229989280469)

('mrsantadora', 0.0030890129632864813)

('robberguy189', 0.0030051476615767735)

האבלואציה עבור הרשת:

ניתן לראות את הרשת כולה ב-layout שנקרא yifan hu -



לאחר סינון וחלוקה לcommunities:

A picture containing doughnut, sky, sitting

Description automatically generated

יש 8 communities וניתן לראות את המשתמשים שיותר הגיבו להם בגדול יותר.

1. Twitter – קיבלנו רשת ממסטרנט במחלקה, אשר מכילה משתמשים שעשו mention למשתמשים אחרים. כל קודקוד הוא יוזר וכל קשת קיימת כאשר יוזר עשה mention ליוזר אחר.

הרשת מורכבת מ8006 קודקודים ומ15418 קשתות.

עשרת הקודקודים עם ה-pageRank הגבוה ביותר הם:

('12', 0.002340343729917841)

('147580943', 0.0011350193066501436)

('15513604', 0.0010451556521229323)

('20733972', 0.0010392536254631693)

('26585095', 0.0009606026404307399)

('280571418', 0.0009538764900086328)

('167536847', 0.0009036483576583567)

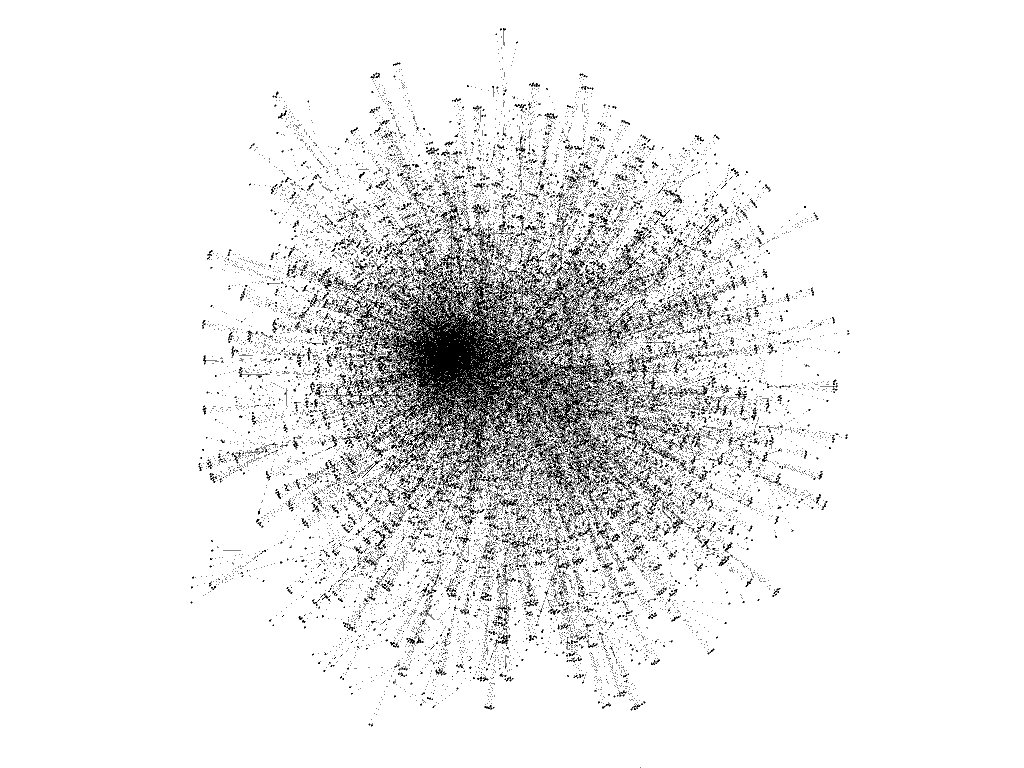
('711114318', 0.0008982911359796004)

('4122226696', 0.0008623496154193781)

('3127990654', 0.0007592674899137956)

האבלואציה עבור הרשת:

ניתן לראות את הרשת כולה ב-layout שנקרא yifan hu -



לאחר סינון וחלוקה לcommunities:

A picture containing sky

Description automatically generated

יש 8 communities וניתן לראות משתמשים שעשו להם הכי הרבה mention בגדול יותר.

1. Wikipedia\_votes – הרשת אשר קיבלנו עם העבודה.

הרשת מורכבת מ7115 קודקודים ומ103689 קשתות.

עשרת הקודקודים עם ה-pageRank הגבוה ביותר הם:

('4037', 0.004607848604934832)

('15', 0.00368018139618871)

('6634', 0.003586612845072666)

('2625', 0.003283828111760979)

('2398', 0.0026086913456084253)

('2470', 0.00252429253191423)

('2237', 0.0024970978239750824)

('4191', 0.00226806883684603)

('7553', 0.002169583332525424)

('5254', 0.002150220879929144)

האבלואציה עבור הרשת:

ניתן לראות את הרשת כולה ב-layout שנקרא yifan hu -

A picture containing animal, invertebrate

Description automatically generated

לאחר סינון וחלוקה ל-communities:

A picture containing sky

Description automatically generated

יש 3 communities וניתן לראות את המשתמשים שהכי הרבה הצביעו להם בגדול יותר.