

ניתוח רשתות חברתיות



פרויקט סיום

לינוי ניסים פור בן לולו
נועה שם טוב

הנתונים נלקחו מאתר SNAP של אוניברסיטת סטנפורד.

הגרף מייצג רשת אמון בין משתמשים שהם כאמור אנונימיים בפלטפורמת Bitcoin OTC

ברשת זו כל משתמש מדרג אחרים בסולם של -10 עד +10 לפי רמת האמון.

מטרת הדירוגים היא למנוע הונאות במסחר אנונימי.

בחרנו לנתח את תת הגרף בעל הSCC הגבוה ביותר

תכונות בסיסיות של הגרף:

• צמתים מיצגים אנשים

◦ מספר הצמתים: 5881

• גרף ממושקל - קשתות מיצגות דירוגים ע"י משקל

◦ מספר הקשתות: 35592

◦ המשקל הנמוך ביותר -10

◦ המשקל הגבוה ביותר הוא 10

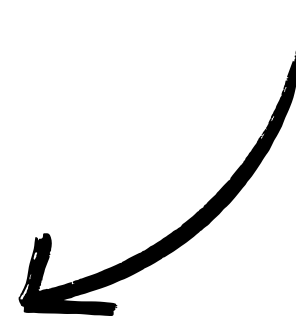
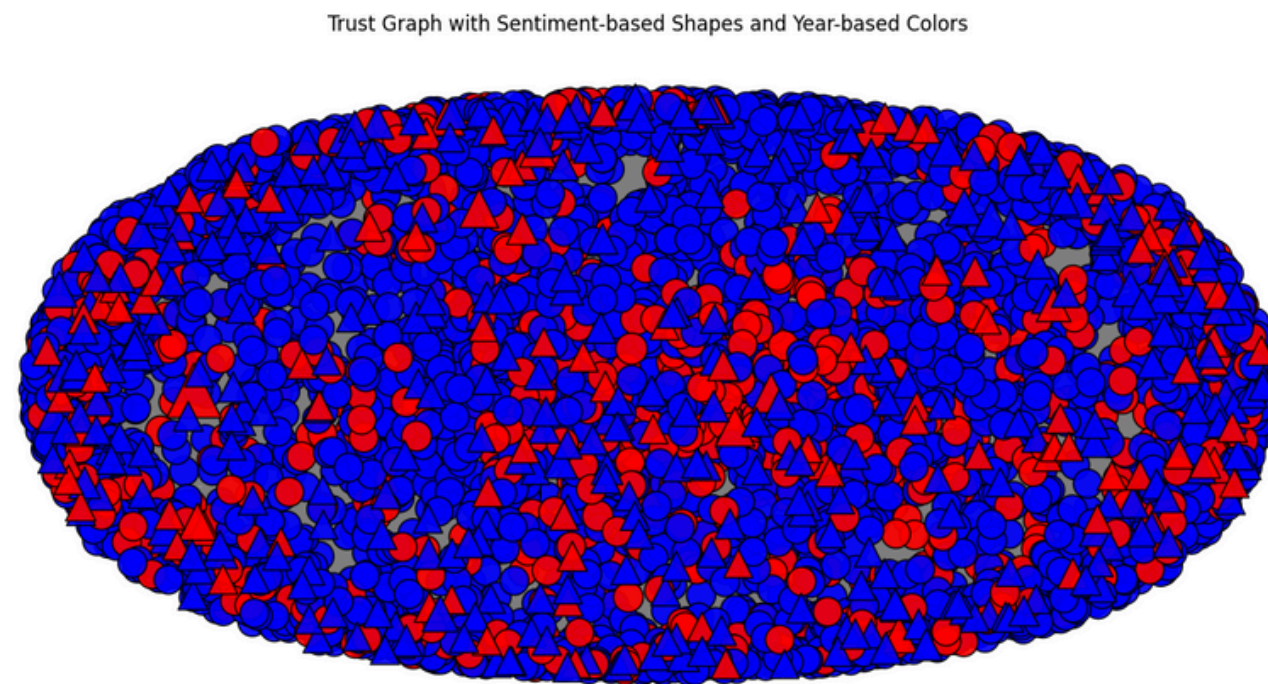
• גרף מכוון

◦ צומת שדורג יהיה בעל קשת נכנסת

◦ צומחת שדירג יהיה בעל קשת יוצאת

• הגרף מכיל 1144 רכיבי קשירות, רובם המוחלט בגודל 1

◦ הSCC הגדול ביותר בגודל 4709 קודקודים ו33461 צלעות



הבחנו בין צמתים שסך דירוגים הוא חיובי לבין כאלו שסך דירוגם שלילי

הפרדנו בין צמתים שדורגו בין השנים 2014-2016 לבין צמתים שדורגו עד 2014

דירוג אחרון שדורגו בין 2014-2016
כמות: 934

דירוג אחרון שדורגו עד 2014
כמות: 3775

סה"כ דורג בחיובי



סה"כ דורג בשלילי



מכיל 1144 SCC
רובם המוחלט גודלם 1
SCC הגדול ביותר גודלו 1144

מכוון

ממושקל

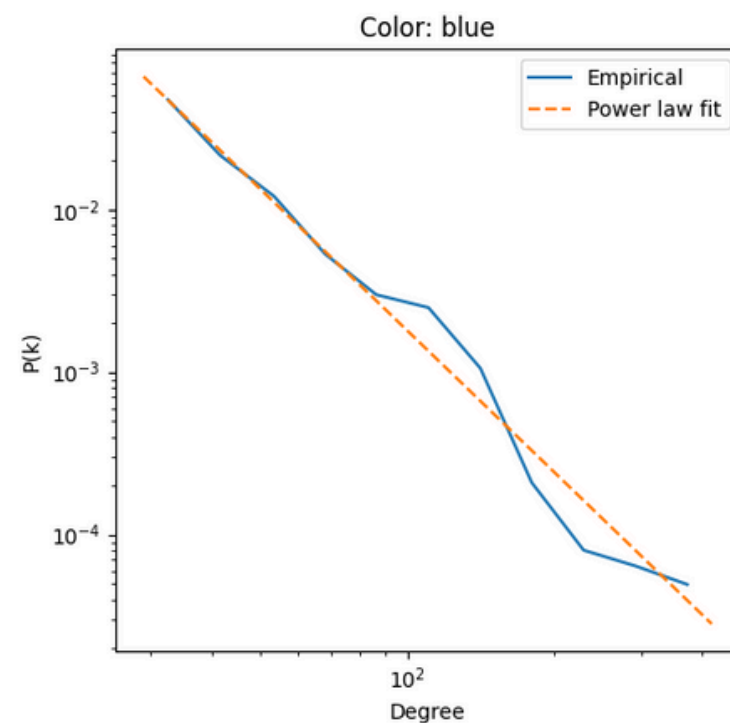
הקוטר הינו 11

העולם הקטן

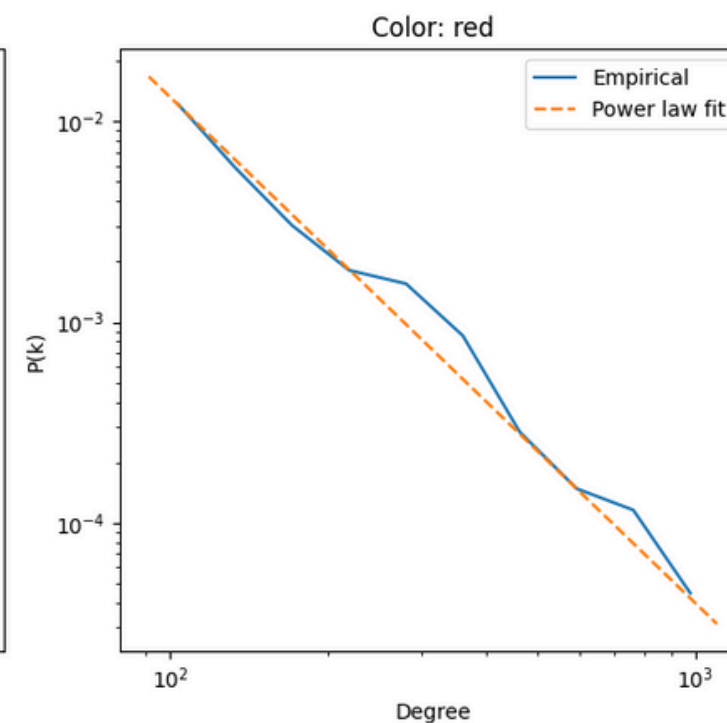
הגרף תומך בתכונת "העולם הקטן", שנבדקה על ידי חישוב אורך המסלול הקצר הממוצע בין כל שני קודקודים. הפונקציה `nx.average_shortest_path_length(G)` החזירה ערך ממוצע של 3.6784.

וקוטר הגרף הינו 11 מצביע על כך שעל אף שהגרף דליל, קיימת קישוריות אפקטיבית מפתיעה: אפילו ברשת לא הדדית, ניתן להגיע מכל משתמש לכל משתמש אחר בלא יותר מ-11 צעדים.

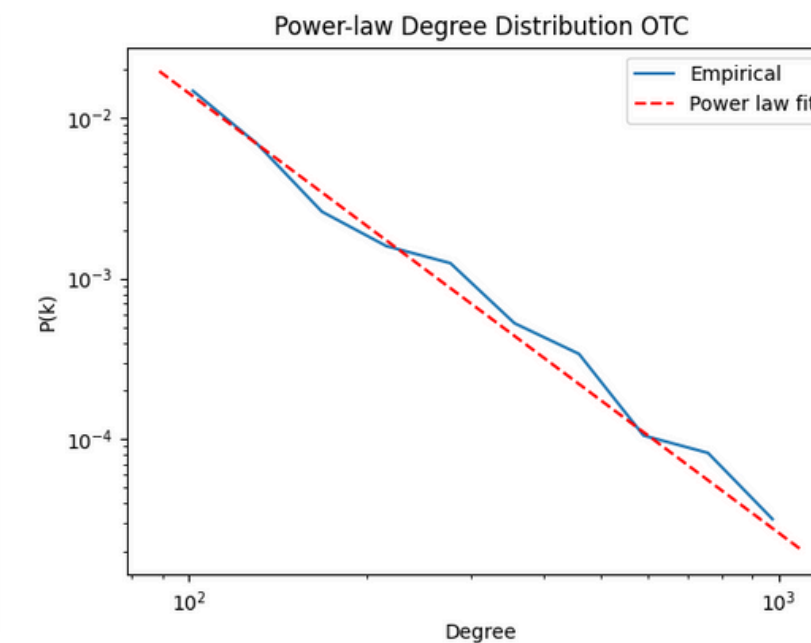
חוק החזקה



Estimated power-law exponent β : 2.91
xmin: 29.0
Number of nodes: 3775

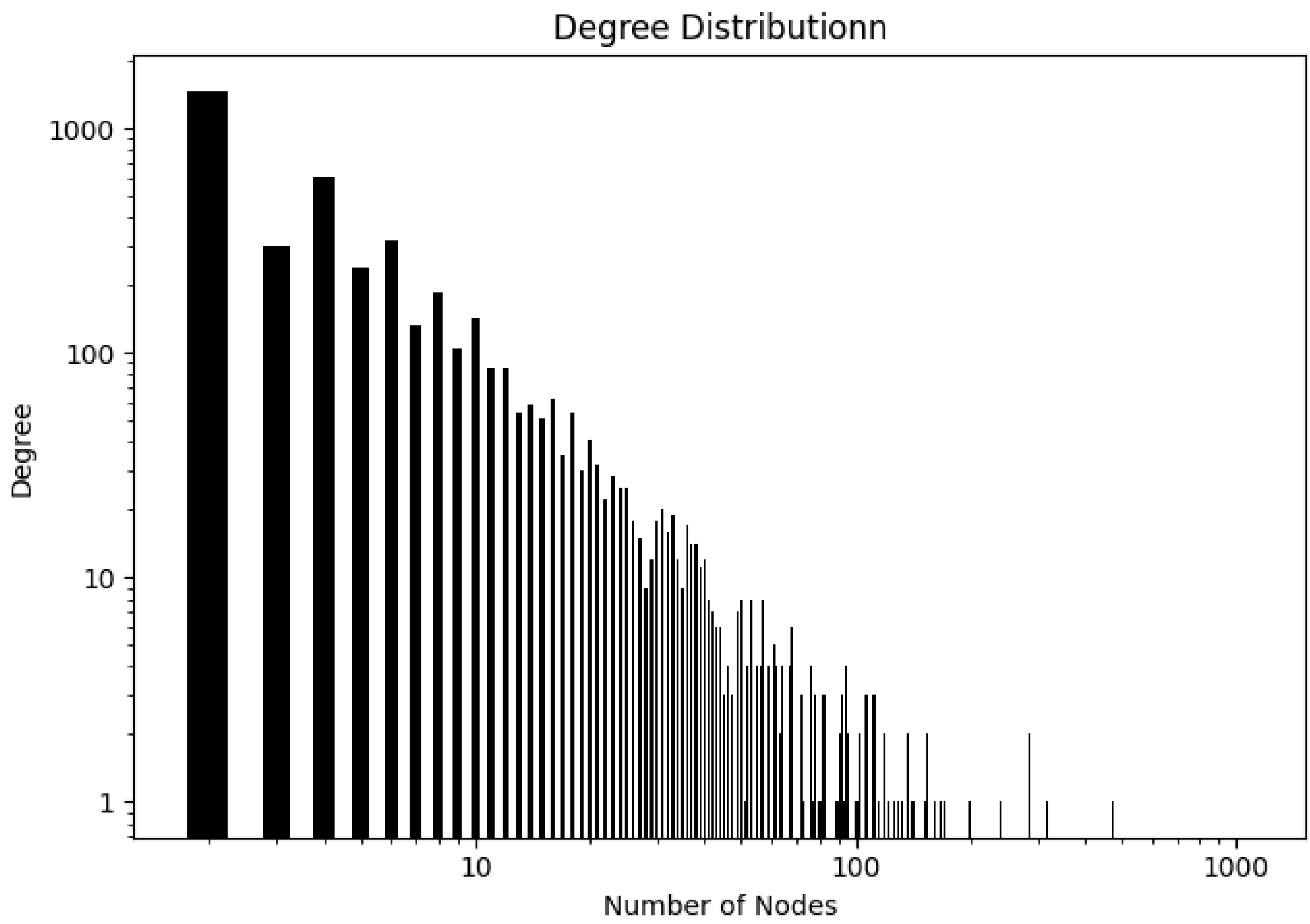


Estimated power-law exponent β : 2.52
xmin: 91.0
Number of nodes: 934



Estimated power-law exponent β : 2.73
xmin: 89.0

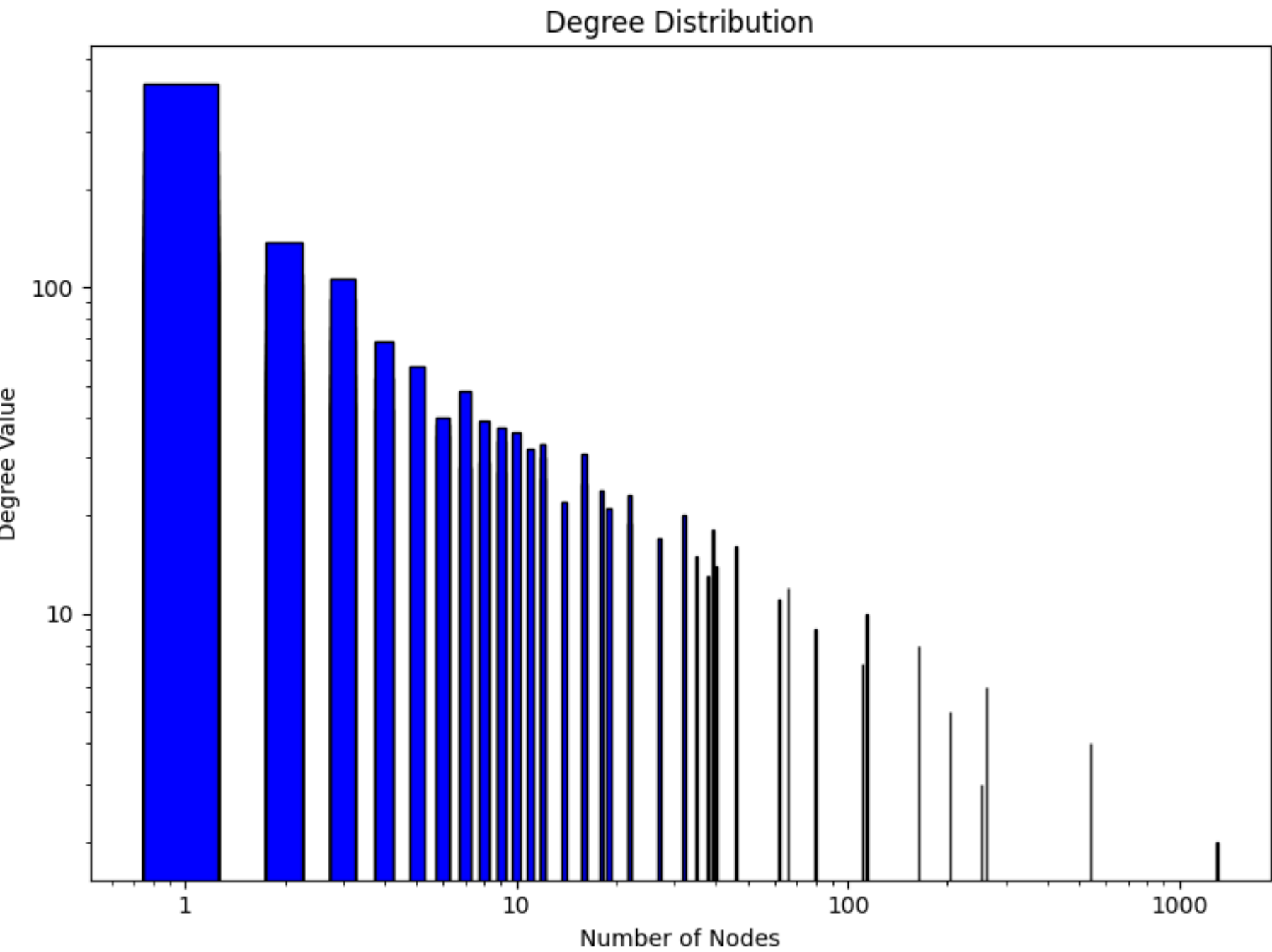
התפלגות הדרגות של הגרף



התפלגות הדרגות של הגרף
לפי צבעים

כחול מייצג קודקודים שדורגו עד שנת 2014

אדום מייצג שנים 2014 - 2016



שאלת המחקר:

מטרתנו לבדוק האם קיימת הטיה בהצבעות בין משתמשים, כלומר – האם אנשים שדורגו באופן חיובי נוטים גם הם לדרג אחרים בצורה חיובית (ולהיפך), מתוך אינטרסים אישיים, או שמא ההצבעות הן אובייקטיביות ומבוססות אך ורק על שיקול דעת עצמאי.

השערת המחקר:

בני אדם הם יצור רציונלי, אך לעיתים מושפעים מהאופן שבו הם נתפסים. לכן, ייתכן כי דירוג אמינות הסוחרים בביטקוין מושפע מדירוגים קודמים, ושהדירוגים אינם אובייקטיביים אלא מבוססים על אינטרסים הדדיים או רגשות תגובתיים.

מודל מתמטי

Degree Centrality

הממצאים מצביעים על כך שחלק מהמשתמשים בעלי in-degree גבוה הם גם בעלי out-degree גבוה, מה שעשוי להעיד על הדדיות בדירוגים.

זה מחזק את ההשערה שהדירוגים אינם תמיד אובייקטיביים, אלא עשויים לשקף יחסים הדדיים או תגובתיות בין משתמשים. השאלה האם יש קשת בין אותם קודקודים פופולרים ואקטיביים ולשם כך נבדוק בעזרת מודל מתמטי נוסף

Out-Degree		In-Degree	
ערך	מספר צומת	ערך	מספר צומת
564	35	534	35
396	2642	412	2642
358	1810	310	1810
283	2028	277	2028
257	905	264	905
251	4172	226	1
239	2125	222	4172
223	7	216	7
207	1	203	4197
203	13	191	13
מדד לאקטיביות אנשים שדירגו הכי הרבה		מדד לפופולריות אנשים שדורגו הכי הרבה	

מודל מתמטי

PageRank

משתמשים משפיעים שזוכים להרבה דירוגים

אז מי הכי אמין?

צמתים שדורגו ע"י אנשים אמינים אחרים יקבלו ציון גבוה

האם ישנה השפעה הדדית בין המשפיענים?

לאחר זיהוי המשתמשים המשפיעים ביותר ברשת (Top-10 לפי PageRank), בדקנו האם קיימת השפעה הדדית ביניהם - כלומר, האם משתמשים אלו דירגו זה את זה באופן ישיר.

ממצאים:

לא נמצאה השפעה הדדית בין המשתמשים המובילים - משתמשים שדירגו משתמשים אחרים ברשימה לא קיבלו מהם דירוג חזרה.

משמעות:

הדבר עשוי להעיד על כך שהשפעתם נובעת מיחסים חד-כיווניים מצד משתמשים אחרים ברשת, ולא ממבנה סגור של קשרים הדדיים בין המשפיעים עצמם

נבדוק באמצעות הומופיליה על מנת לוודא את התוצאה הזו

PageRank	
מספר צומת	ערך
2642	0.0179
35	0.0143
1	0.0106
7	0.0099
1810	0.0082
4172	0.0078
2028	0.0073
1018	0.0068
905	0.0065
2125	0.0057

מהי הומופיליה?

הומופיליה היא הנטייה של משתמשים ליצור קשרים עם אחרים שדומים להם הדמיון יכול להתבסס על: מאפיינים אישיים, התנהגות, עמדות

בניתוח שלנו התמקדנו בהומופיליה התנהגותית: סיווגנו את המשתמשים לשלוש קבוצות לפי ממוצע הדירוגים שהם נותנים:

- Positive raters – ממוצע דירוגים מעל 2
- Negative raters – ממוצע מתחת ל-2
- Neutral – כל השאר

בדקנו האם משתמשים נוטים לדרג אחרים מאותה קבוצת התנהגות. אם כן – מדובר בהומופיליה, מה שיכול להעיד על הטיה סובייקטיבית ולא על שיפוט אובייקטיבי.

לאחר ניתוח הגרף יצא שדירוגים חיוביים ושליילים מתבצעים בעיקר מחוץ לקבוצה

לכן, נראה כי משתמשים לא בהכרח מעדיפים לדרג משתמשים עם סגנון דירוג דומה, אלא כנראה פועלים לפי שיקולים פרטניים (למשל אמון אישי, עסקה ספציפית, ניסיון קודם).

מסקנות

מתוצאות הניתוח שביצענו על הדאטה עולה כי הרשת אובייקטיבית וההשערה לפיה אדם מושפע ומונע מאינטרסים כרגע נדחתה

עם זאת, אפשר לחשוב על כיוונים נוספים למחקר:

השפעה בין מדרגים שמדרגים אחד את השני:

בדאטה יש כ־14,000 מקרים של זוגות (מעגלים בגודל 2), שבהם אדם דירג מישהו, והוא דירג אותו בחזרה. אפשר לבדוק האם הדירוג השני (שקרה מאוחר יותר) הושפע מהדירוג הראשון. כלומר, האם הם עשו זאת בצורה אובייקטיבית או סובייקטיבית?

חוסר סימטריה ודפוסים בדירוגים:

אפשר לבדוק האם קיימים דפוסים עקביים אצל מדרגים מסוימים - לדוגמה, אנשים שנוטים לתת ציונים גבוהים יותר או נמוכים יותר מהממוצע באופן שיטתי. דפוסים כאלה עשויים להעיד על הטיות קבועות של חלק מהמשתמשים.

אמינות לפי יחס דירוגים:

כיוון נוסף למחקר הוא לבדוק האם במערכת שבה יש אנשים שמדרגים מעט אך זוכים להרבה דירוגים מאחרים, ניתן להניח שאין אינטרס אישי או מוטיבציה להטות את הדירוגים. חקר האסימטריה בין כמות הדירוגים שמישהו נותן לבין כמות הדירוגים שהוא מקבל יכול להצביע על אמינות ויעילות של המערכת כולה.