**המחלקה להנדסת תוכנה**

**פרויקט גמר – ה'תשפ"ד**

**מודלים מקריים עבור רשתות חישוב מקבילי**

**Random Models For Parallel Computing Networks**

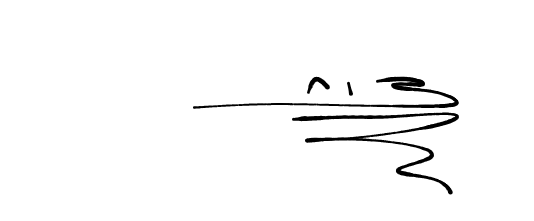
**מאת**

**עופרי אברמוביץ'**

**209073741**

**יאיר רחמני**

**207725789**



**מנחה אקדמי/ת: ד״ר צור לוריא אישור: תאריך: 16/04/24**

מערכות ניהול הפרויקט:

|  |  |  |
| --- | --- | --- |
| # | מערכת | מיקום |
| 1 | מאגר קוד | <https://github.com/ofriaab/random_models_final_project> |
| 2 | קישור ליומן | <https://github.com/ofriaab/random_models_final_project/issues> |
| 3 | קישור לסרטון דוח אלפא | <https://drive.google.com/file/d/1FLVlsyTEB-0S6WWnmqw7HKVWWTDAEMNs/view?usp=sharing> |

מידע נוסף (מחקו את המיותר)

|  |  |
| --- | --- |
| סוג הפרויקט | מחקרי ממרצה במכללה בשיתוף הייטק |
| פרויקט ממשיך | זה פרוייקט חדש |

**תוכן עניינים**

**נאום המעלית**1

[**מבוא**](#_30j0zll) 1

**תיאור הבעיה** 2

[דרישות ואפיון הבעיה](#_3znysh7) 2

[הבעיה מבחינת הנדסת תוכנה](#_2et92p0) 2

[**סקירת עבודות דומות בספרות והשוואה**](#_tyjcwt) 3

[**תיאור הפתרון**](#_3dy6vkm) 4

[**מה עשינו עד כה?**](#_4d34og8) 5

**מסקנות**5

[**נספחים**](#_17dp8vu) 6

גרפים6

[תכנון הפרויקט – ברזולוציה של שבועיים](#_lnxbz9) 8

[רשימת ספרות (ביבליוגרפיה)](#_35nkun2) 8

# נאום המעלית

תחום הרשתות המורכבות הינו תחום גדול ומתפתח, המשמש להבנה ומידול של מערכות יחסים והתנהגויות מורכבות המצויות במערכות מכוונות. ישנם אלגוריתמים שונים והיוריסטיקות עבור מודלים אקראיים ליצירת גרפים מכוונים, אך לא קיימים מודלים ליצירת גרפים מכוונים המדמים נתונים מהעולם האמיתי. בפרויקט זה אנו נתמקד בגרפי תלויות מכוונים, על בסיס גרפי תלויות המייצגים אלגוריתמים מקבילים של חברת mobileye אנו נחקור את תכונות גרפים אלו במטרה לפתח מודל אקראי ליצירת גרפים מכוונים שיתבסס על תכונות אלו וידמה גרפים מהעולם האמיתי.

# מבוא

בתחום ההולך ומתרחב של רשתות מורכבות, האתגר של בניית מודלים המבוססים על נתונים מהעולם האמיתי מהווה משימה גדולה ומשמעותית. גרפים ורשתות שנגזרו מנתונים ממשיים מכילים את המפתח להבנת התנהגותן של מערכות מורכבות, אך פיתוח מודלים מדויקים למטרה זו נותר אתגר מתמשך. אתגר זה בולט במיוחד בתחום של גרפים מכוונים, בהם היחסים בין צמתים משקפים תלות מהעולם האמיתי.

גרפים מכוונים משמשים ככלי יסוד להבנת התנהגותן של מערכות שונות, החל מרשתות חברתיות ועד לתשתיות תחבורה. עם זאת, למרות נוכחותם בכל מקום, קיים מחסור בולט של ספרות המתייחסת לגרפים מכוונים באופן מקיף. פער זה בידע מעכב את היכולת שלנו לפתח מודלים מדויקים שתופסים את נבכי התלות בעולם האמיתי.

בהקשר זה, תחום האלגוריתמים המקבילים מתגלה כמוקד עניין. אלגוריתמים מקבילים, שנועדו לרתום את הכוח של מספר מעבדים, ממלאים תפקיד מרכזי במערכות מחשוב מודרניות. עם זאת, הבנת ההתנהגות של אלגוריתמים מקבילים דורשת תובנה עמוקה יותר לגבי גרפי התלויות העומדים בבסיס פעולתם. גרפי תלות אלו, המייצגים את היחסים בין משימות באלגוריתם מקביל, מספקים תובנות חיוניות לגבי זרימת החישוב והקצאת המשאבים.

תזמון משימות בצורה אופטימלית בין מעבדים שונים זוהי בעיה NP קשה ומציאת פתרון יעיל לכך מהווה אתגר גדול. יתרה מכך, המגוון של סוגי המעבדים מסבך עוד יותר את תהליך התזמון, שכן למעבדים שונים יש יכולות ואילוצים משתנים. בתחום של מערכות בזמן אמת, שבהן מגבלות התזמון מחמירות, הצורך באלגוריתמי תזמון יעילים הופך בולט עוד יותר.

גישה נפוצה אחת לתזמון משימות בחישוב מקביל כוללת סיווג של משימות לשתי קטגוריות: משימות עצמאיות ומשימות תלויות. עבור משימות עצמאיות, ההתמקדות היא לעתים קרובות בהשלמת חישובים כבדים תחילה (LTF), תוך מינוף כוח העיבוד הזמין כדי להתמודד עם משימות עתירות משאבים מראש. לעומת זאת, עבור משימות תלויות מאוד, הדגש עובר לעבר השלמת משימות קלות תחילה (STF), תוך מתן עדיפות לפתרון התלות כדי למנוע צווארי בקבוק.

עם זאת, הנוף הקיים של אלגוריתמי תזמון והיוריסטיקה אינו יכול להתמודד עם המורכבות של תרחישים בעולם האמיתי. בעוד שהיוריסטיקות מציעות גישה פרקטיות להתמודדות עם בעיות NP-שלמות בתרחישי זמן אמת, הן לרוב אינן מצליחות להסביר את הניואנסים של נתונים מהעולם האמיתי ואת המורכבויות של חישוב מקביל. בפועל, לאלגוריתמים ולמודלים הקיימים אין הרבה קשר לנתונים מהעולם האמיתי, וכתוצאה מכך מניבים ביצועים לא אופטימליים וחוסר יעילות.

לאור אתגרים אלו, הפרויקט שלנו שואף לגשר על הפער בין המודלים הקיימים למורכבויות בעולם האמיתי. על ידי פיתוח מודלים של גרפים מכוונים הלוכדים את התלויות באלגוריתמים מקבילים - על בסיס תובנות ממערכות בעולם האמיתי ובפרט נתוני אלגוריתמים מקבילים של חברת Mobileye. אנו שואפים לחולל מהפכה בתחום התזמון והחישוב המקביל. המטרה שלנו היא לספק בסיס חזק יותר לתזמון אלגוריתמים ורשתות מחשוב מקבילות, ובסופו של דבר לשפר את היעילות והביצועים של מערכות מורכבות. באמצעות ניתוח מדוקדק ופיתוח מודלים, אנו שואפים להסיק תובנות חדשות, לשפר יכולות חיזוי ולסלול את הדרך למערכות יעילות ואפקטיביות יותר בתחום הרשתות המורכבות.

# תיאור הבעיה

דרישות ואפיון הבעיה

כיום קיימים אלגוריתמים לפיתוח מודלים מקריים עבור רשתות חישוב מקבילים אך אלגוריתמים אלו אינם מדמים או מסתמכים על תכונות מהעולם האמיתי. המודלים ברובם שרירותיים ופשוטים ואינם מנצלים את היתרונות בשימוש באילוצי העולם האמיתי על מנת לשפר ולדייק את המודלים הללו. הבנת ההתנהגות של אלגוריתמים מקבילים היא חיונית באופטימיזציה של ביצועים והבטחת אמינות, במיוחד במערכות מורכבות כמו מערכת הנהיגה האוטומטית של Mobileye. היעדר מודלים לגרפי תלויות מכוונים המייצגים אלגוריתמים מקבילים מציב אתגר עבור חוקרים ומהנדסים. פער זה מעכב ניתוחים וסימולציות מקיפות החיוניות לאימות אלגוריתמים ושיפור ביצועים. טיפול בבעיה זו דורש חקירה של גרפי התלויות של האלגוריתמים של חברת Mobileye ופיתוח שלאחר מכן של מודלים סטוכסטיים בהתבסס על התכונות שמצאנו.

## 

## הבעיה מבחינת הנדסת תוכנה

מנקודת מבט של הנדסת תוכנה, היעדר מודלים ליצירת גרפים מכוונים עבור אלגוריתמים מקבילים מעכב את הפיתוח, הבדיקה והאימות של אלגוריתמים. מערכות בעולם האמיתי, כמו אלגוריתם הנהיגה האוטומטי של Mobileye, מסתמכות במידה רבה על עיבוד מקבילי, מה שהופך את דימוי והבנת התכונות של גרפים אלו לחיוני. היעדר מודלים מתאימים מגביל את היכולת להעריך ביצועי אלגוריתמים תחת תרחישים שונים, מה שמפריע לתהליך הפיתוח האיטרטיבי. הפרויקט שלנו שואף לגשר על פער זה על ידי שימוש בעקרונות הנדסת תוכנה לחקירת גרפי תלויות מכוונים מהעולם האמיתי בהתבסס על הדאטה שקיבלנו מחברת מובילאיי ויצירת מודלים יעילים, המאפשרים ניתוח מקיף ושיפור של אלגוריתמים מקבילים. האתגר טמון ביצירת עומסי עבודה סינתטיים שיתוארו באמצעות גרפים כדי לאמת ולנתח אלגוריתמי תזמון יעילים. לשיטות הקיימות אין כלי סטנדרטי ורב-תכליתי ליצירת גרפי משימות בעלי מאפיינים מהעולם האמיתי. מתן מענה לבעיה זו יאפשר למפתחי אלגוריתמים להעריך במדויק את הפתרונות שלהם תחת תרחישים שונים וימנע עיצוב אלגוריתמים לא אופטימליים בעלי ישימות מוגבלת בעולם האמיתי.

הפרויקט שלנו מציג מספר אתגרים. ראשית, תכנון אלגוריתמים יעילים להפקת מודלים של גרפים מכוונים והדמיית רשתות מחשוב מקבילות הוא בעל חשיבות עליונה. אלגוריתמים אלה חייבים להיות ניתנים להרחבה ולאופטימיזציה לביצועים כדי להתמודד עם מערכי נתונים גדולים ביעילות. טיפול יעיל בנתונים הוא גם חיוני, מחייב יישום של מנגנונים חזקים לניהול נתוני קלט ועיבוד כמויות גדולות של נתונים ביעילות. אימות מודל הוא היבט מרכזי נוסף, הכולל השוואה של תוצאות סימולציה עם נתונים מהעולם האמיתי כדי להבטיח את הדיוק והיעילות של המודלים. יתר על כן, תכנון ארכיטקטורת תוכנה גמישה וניתנת להרחבה חיוני כדי להתאים לשיפורים ושינויים עתידיים. מאמצי אופטימיזציית הביצועים יתמקדו במזעור תקורה חישובית ושיפור היעילות, במיוחד כאשר עוסקים במערכי נתונים בקנה מידה גדול.

# סקירת עבודות דומות בספרות והשוואה

כדי להתחיל לעסוק בשאלת המחקר חיפשנו האם ישנם מאמרים העוסקים בנושא יצירת גרפי תלויות מכוונים מקריים עבור בעיות תזמון המבוססים על נתונים מהעולם האמיתי. נוכחנו לגלות כי יש מספר רב של מאמרים שמתעסקים בנושא גרפי תלויות מקריים עבור בעיות תזמון, אך אינם מבוססים על דאטה אמיתי של אלגוריתמי תזמון ולכן אינם עונים על שאלת המחקר שלנו. בסקירה זו נתייחס לשניים מתוך המאמרים ובנוסף, נסיק מסקנות מתוך ספר רחב היקף בנושא רשתות מורכבות אשר לא נוגע כלל בגרפי תלויות ובעיות תזמון .

המאמרים "Random graph generation for scheduling simulations"  ו"A Comparison of Random Task Graph Generation Methods for Scheduling Problems" מתעמקים באתגרים ובמתודולוגיות סביב יצירת גרפי תלויות מקריים עבור בעיות תזמון במערכות מבוזרות ומקביליות  תוך שימוש באלגוריתמיים מקריים ידועים כגון : The G(n, p) method ו  Layer-by-Layer method הידועים זה מכבר  שנים רבות .

במאמר "Random graph generation for scheduling simulations" , הכותבים מציגים את GGen, כלי שמטרתו לספק יישום אחיד וסטנדרטי של שיטות יצירת גרפי תלויות קלאסיות. הם מדגישים את החשיבות של הבנת המאפיינים של הגרפים שנוצרו ועל השפעתם על אלגוריתמי תזמון, מתוך הכרה שבחירה לא נכונה של אלגוריתם יצירה מקרי עלולה להוביל לתוצאות לא מדויקות דיו. בעוד ש-GGen משתמש באלגוריתמים ידועים של יצירה מקרית, הוא נופל במתן מענה לצורך ביצירת גרפים לא מקריים, שהוא חיוני בשל האילוצים והמאפיינים של העולם האמיתי הטמונים בבעיות תזמון.

המאמר A Comparison of Random Task Graph Generation Methods for Scheduling Problems"" מחזק את המשמעות של יצירת גרפי משימות עם מאפיינים רלוונטיים וללא הטיה. המאמר מזהה מאפייני מפתח, כגון מסה, המכמתים את המאפיינים הספציפיים למקרים קשים, ומסייעים בבחירה של מקרים מאתגרים לאימות אלגוריתם. עם זאת, בדומה למאמר הראשון, הוא מתמקד בעיקר בניתוח שיטות קיימות ליצירת גרפים אקראיים מבלי לחקור חלופות המותאמות להבנת המורכבות של תרחישי תזמון בעולם האמיתי.

בנוסף הספר " Complex Networks: Principles, Methods and Applications 1"  סוקר בהרחבה שלל נושאים ברשתות מורכבות , אך אינו מציין כלל את נושא בעיות התזמון ושימוש בגרפי תלויות . בעוד שרשתות מורכבות משמשות מסגרת תיאורטית להבנת מערכות שונות, כולל מערכות תזמון, היעדר העיסוק בנושא DAGs ובעיות תזמון מדגיש את הצורך במחקר מיוחד בתחום זה. ספר זה מדגיש את השוני בין ההיקף הכללי של תורת הרשת המורכבת לבין הדרישות הספציפיות של בעיות תזמון, ומדגיש עוד יותר את הצורך בגישות מותאמות בתחום זה.

לסיכום, המאמרים מספקים תובנות חשובות לגבי האתגרים והמתודולוגיות של יצירת גרפי משימות עבור אלגוריתמי תזמון, בעוד שהספר מדגיש את היעדר העיסוק בנושא והחשיבות של מחקר נוסף של רשתות מורכבות כחלק מפתרון בעיות התזמון השונות בעולם האמיתי. מחקר זה נועד לשים דגש על חקירת גרפי תלויות ותכונות של רשתות מורכבות על מנת לפתח שיטות יצירת גרפים לא מקריים עבור בעיות התזמון המורכבות של העולם האמיתי .

# תיאור הפתרון

הפתרון שלנו לבעיה כולל שלושה שלבים עיקריים, שכל אחד מהם נועד לתרום להבנה מקיפה של התלויות הבסיסיות והשפעתן על ביצועי האלגוריתמים.

ראשית, אנו מתחילים בניתוח יסודי של גרפי התלויות המופקים מאלגוריתמי Mobileye. שלב זה כולל התעמקות במורכבות האלגוריתמים, וזיהוי תכונות ומאפיינים מרכזיים של הגרפים המכוונים המייצגים תלות במשימה. על ידי בחינת התלויות באלגוריתמים, אנו שואפים להבחין בהיבטים קריטיים המשפיעים על הביצועים. המחקר הבסיסי הזה מאפשר לנו להסיק מסקנות לגבי המבנה של גרפים מהעולם האמיתי.

בשלב השני, אנו נעסוק בפיתוח מודלים אקראיים המותאמים ליצירת גרפי תלויות מכוונים על בסיס ניתוח האלגוריתמים שבצענו. תוך מינוף תובנות שנאספו מהניתוח הגרפי שלנו, אנו ניצור מודל אקראי המסוגל לייצר גרפי תלויות מכוונים המשקפים אלגוריתמים מקביליים מהעולם האמיתי. המיקוד שלנו טמון ביצירת תהליך פשוט אך יעיל להפקת גרפים אקראיים שמכילים את המורכבויות שגילינו מחקר אלגוריתמי Mobileye. באמצעות תשומת לב קפדנית לפרטים, אנו שואפים להבטיח שהמודלים יקלטו במדויק את הניואנסים של תלויות בעולם האמיתי, ובכך ישפרו את אמינות הסימולציות שלנו.

לאחר יצירת מודלים אקראיים ליצירת גרפי תלויות מכוונים, אנו נבצע סימולציה של אלגוריתמי תזמון על הגרפים שיצרנו.אנו נעריך את ביצועי אלגוריתמי התזמון על הגרפים שנוצרו. על ידי הכפפת האלגוריתמים לבדיקה וניתוח קפדניים, אנו נשיג תובנות חשובות לגבי ההשפעה של התלויות בין המשימות על ביצוע אלגוריתמים מקבילים. תהליך איטרטיבי זה יאפשר לנו לחדד את ההבנה שלנו לגבי ביצועים אלגוריתמיים בהקשר של תלויות מן העולם האמיתי, ולסלול את הדרך לאסטרטגיות יעילות ואופטימיזציה.

מבחינת ארכיטקטורת המערכת והכלים העיקריים, אנו נשתמש בPython כשפת התכנות העיקרית לניתוח וסימולציה של אלגוריתמים, בנוסף נשתמש בספריות ניתוח רשת לחקר הגרפים.

# מה עשינו עד כה?

* ניתוח גרפי תלויות של חברת mobileye- תחילה עבדנו עם גרפי תלויות פשוטים המיוצגים על ידי קבצי deps.

תכונות מרכזיות שנחקרו כללו:

* התפלגות מספר השכנים הנכנסים והיוצאים מכל צומת (נספחים 1,2)
* התפלגות רוחב הגרפים (נספח 3)
* ניתוח גרפי תלויות של חברת mobileye הכוללים זמני ריצה של המשימות המיוצגות בגרף. בשלב זה בחנו את זמן הריצה של הנתיב הקריטי אל מול זמן הריצה הכולל של האלגוריתם.  
  (נספח 4)  
    
  משימות לביצוע בהמשך :
* המשך ניתוח של גרפי התלויות על ידי בחינת תכונות ומדדים נוספים, במטרה להבנה מקיפה יותר של תלות אלגוריתמית.
* השוואת תכונות הגרפים שקבלנו מהניתוח אל מול תכונות הגרפים הקיימים היום עליהם עברנו בסקירה הספרותית.
* פיתוח מודלים סטוכסטיים המבוססים על ממצאי הניתוח, המשמשים ככלי יסוד להדמיית רשתות מחשוב מקבילויות.
* ביצוע סימולציה של ריצת אלגוריתמי תזמון על גרפי התלויות שיצרנו כדי לאפשר הערכת ביצועים.

**מסקנות -**

* מסקנות מנספחים 1,2,3:

בניתוח מספר הדרגות הנכנסות והיוצאות ורוחב הגרפים קיבלנו מאפיינים מבניים של הגרפים והמורכבות והגודל הכוללים של הגרפים, ממצאים אלו יסייעו לנו בבניית מודלים אקראיים לגרפים המבוססים על נתונים מהעולם האמיתי.   
  
בנוסף, ממצאים אלו ישמשו אותנו כאמצעי לאימות המודלים שלנו. על ידי השוואת התפלגות הדרגות והרוחבים בנתוני העולם האמיתי עם אלו שנוצרו על ידי המודלים שלנו, נוכל להעריך את הדיוק והנאמנות של המודלים שלנו. תהליך אימות זה יבטיח שהמודלים שלנו יקלטו ביעילות את המאפיינים המבניים והתלויות שנצפו בתרחישים בעולם האמיתי.

* מסקנות מנספח 4:  
  בהתבסס על הממצאים שלנו מניתוח זמני הריצה של הגרפים, אנו יכולים להסיק מספר מסקנות משמעותיות לגבי הקשר בין זמן הריצה של הנתיב הקריטי לבין זמן הריצה הכולל של הגרפים.

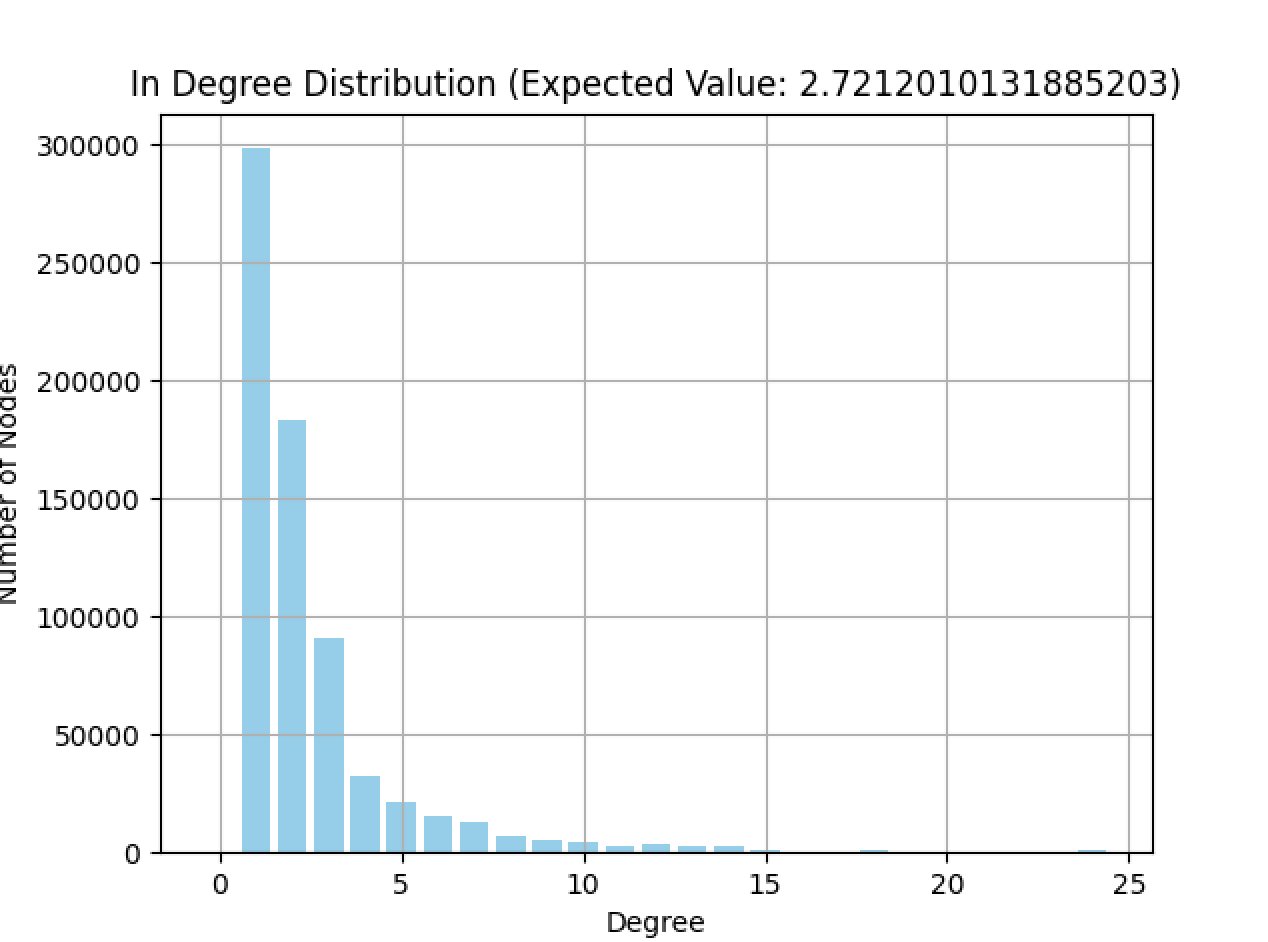
זמן ריצת הנתיב הקריטי אל מול זמן הריצה הכולל כולל (CPD/TD): היחס הממוצע הוא כ-0.1719, אנו רואים שהנתיב הקריטי מהווה כ-17.19% מזמן הריצה הכולל.

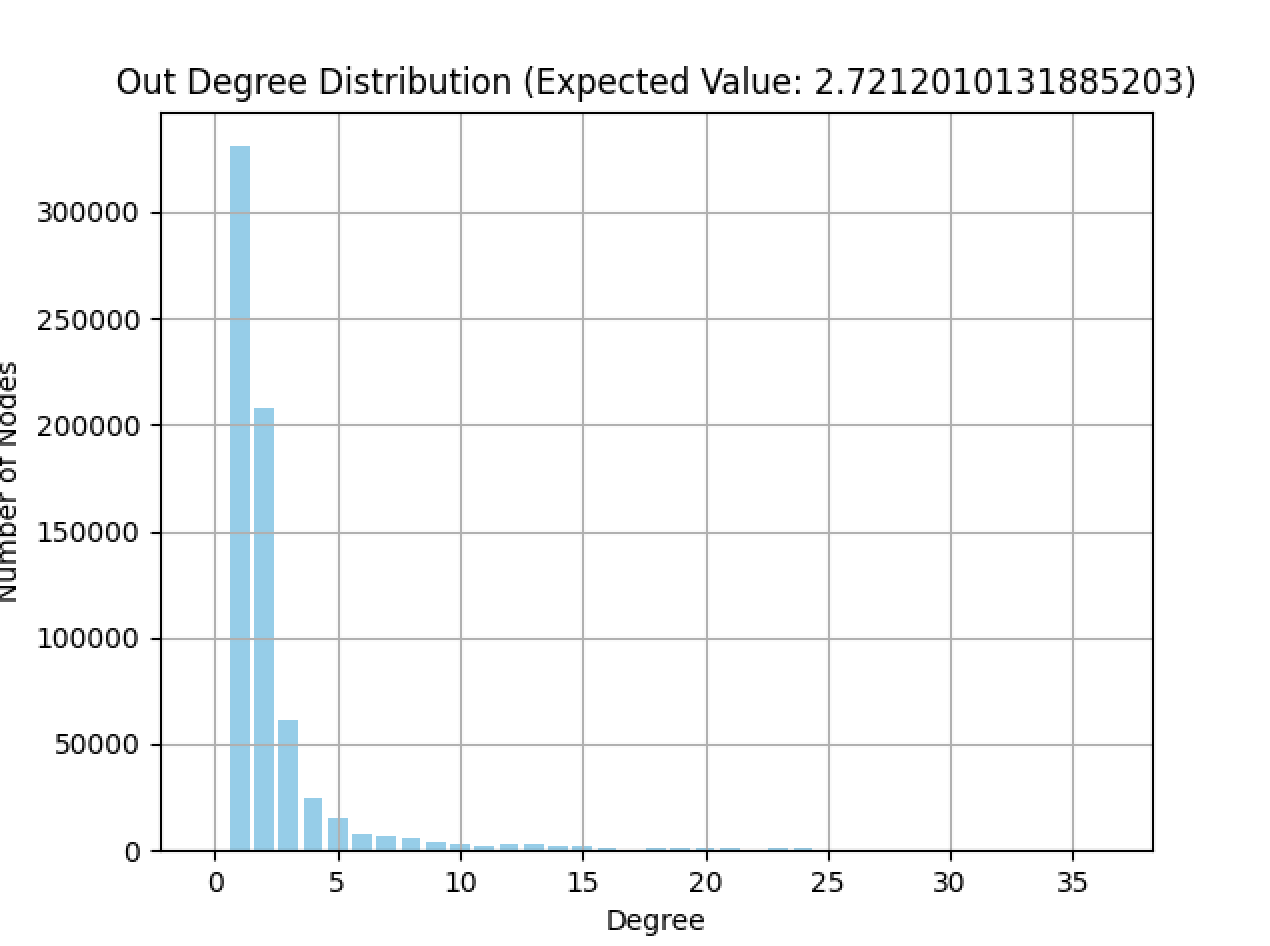
מספר ממוצע של הצמתים (משימות) בנתיב הקריטי הוא כ79ת ממוצע הצמתים הכולל בגרף הוא כ-7,121 צמתים.

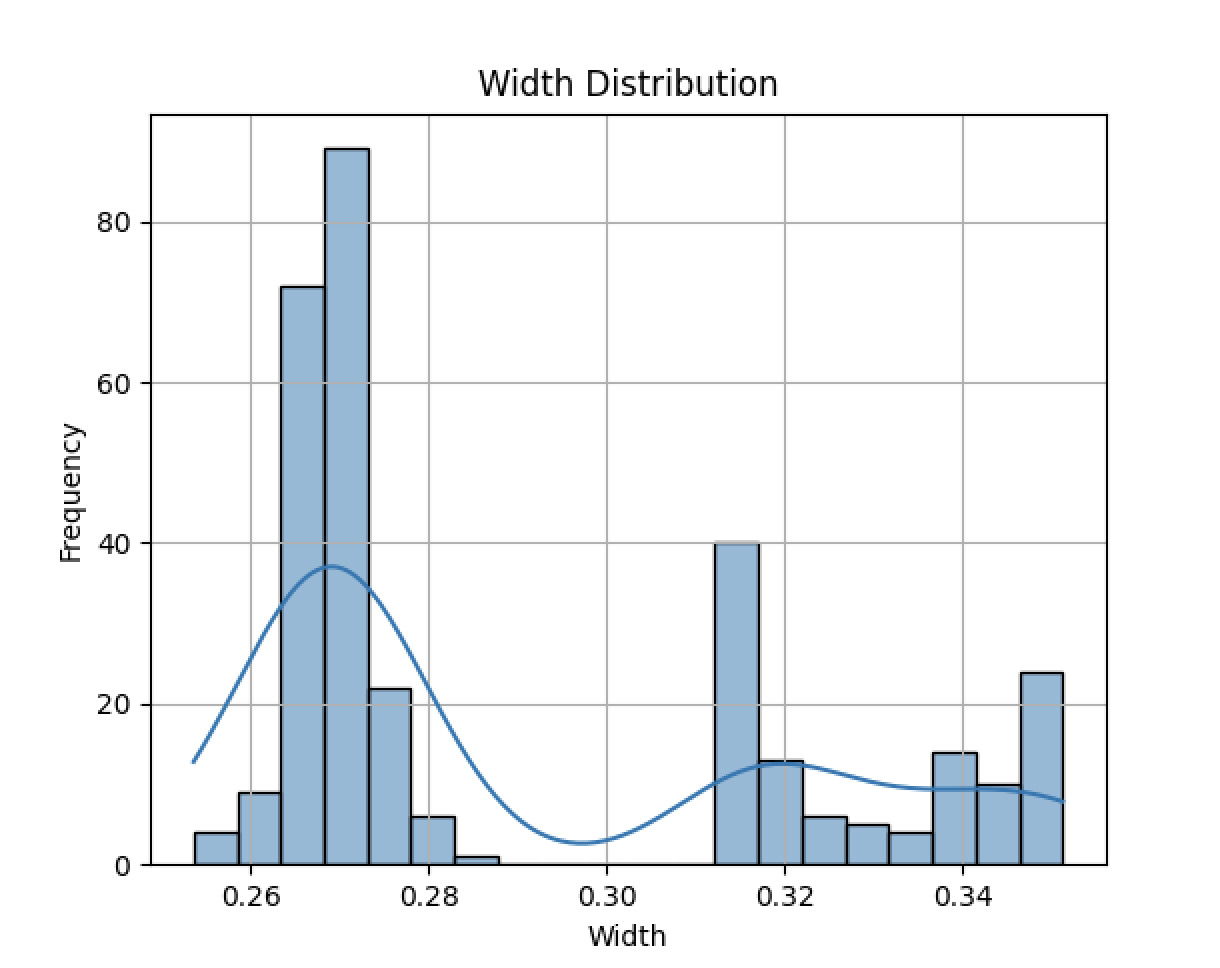
היחס בין מספר הצמתים בנתיב הקריטי למספר הכולל של הצמתים בגרף מספק תובנות לגבי הגודל היחסי של הנתיב הקריטי בתוך הגרף. עם יחס גודל נתיב קריטי ממוצע של כ-0.0113, אנו רואים שהנתיב הקריטי מהווה כ-1.13% מסך הצמתים בגרף.

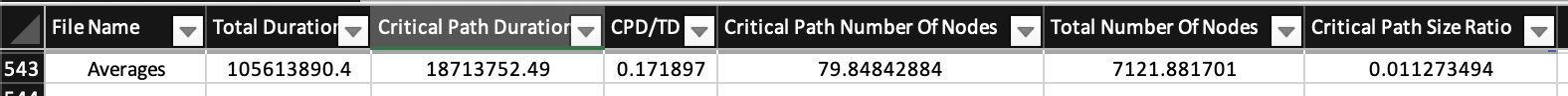
יחס ה- CPD/TD מציג הזדמנות לאופטימיזציה. המשימות שאינן משתייכות לנתיב הקריטי תורמות לרוב זמן הריצה, התמקדות באופטימיזציה שלהן יכולה להפחית משמעותית את זמן הריצה הכולל. על ידי זיהוי חוסר יעילות ומינוף מקביליות, אנו יכולים לשפר את הביצועים האלגוריתמיים ולשפר את היעילות בתוך רשתות מחשוב מקבילות. ממצא זה מתיישב עם מטרת הפרויקט שלנו לפתח מודלים הלוכדים במדויק תלויות והתנהגויות בעולם האמיתי, ויובילו לאלגוריתמים מקבילים ורשתות מחשוב יעילים יותר.

# נספחים

נספח 1:****

נספח 2:****

נספח 3:****

נספח 4:

## תכנון הפרויקט – ברזולוציה של שבועיים

|  |  |
| --- | --- |
| 30.4 | ניתוח סטטיסטי של המסקנות |
| 7.5 | בניית מודל מקרי |
|  | תכנון |
|  | פיתוח |
| 21.5 | בדיקות |
| בהמשך | בניית מערכת סימולציה |
|  | הרצת סימולציות על הדאטה האמיתי והמלאכותי |
|  | השוואה וניתוח של הנתונים |

## רשימת ספרות (ביבליוגרפיה)

1. [//hal.science/hal-00471255/document](https://hal.science/hal-00471255/document)
2. <https://arxiv.org/pdf/1902.05808.pdf>
3. Complex Networks: Principles, Methods and Applications 1