

Biological computation – HW2

רון ימין 315252833 עומרי צדוק 208726802

הסבר על הקוד:

קוד שאלה 1:

בשביל לפתור את הבעיה בשאלה הזו השתמשנו בספריית networkx, ספרייה שמאפשרת יצירה והשוואה של גרפים, בפרט השתמשנו ב: DiGraph – בשביל יצירת גרפים מכוונים is_weakly_connected – בשביל לבחור רק גרפים שמתאימים לשאלה, (הגרפים צריכים להיות מחוברים)

is_isomorphic – בשביל להימנע למספור את אותם גרפים מספר פעמים

בנוסף, בשביל לעבור על כל הצירופים האפשריים לגרפים בגודל n השתמשנו ב-combinations תחת הספרייה itertools.

התוכנית עובדת בלולאה מ- $n=1$ עד $n=4$ (מכיוון שבערכי n האלו זמן הריצה מהיר בניגוד לערכי n גדולים יותר), כך בשכל פעם, אנחנו מייצרים את כל הקשתות האפשריות, לאחר מכן אנחנו עוברים על כל הקומבינציות האפשריות של הקשתות (באמצעות combinations) כלומר אנו עוברים על $2^{n(n-1)}$ גרפים.

עבור כל קומבינציה אנחנו בונים גרף מכוון, אנחנו בודקים האם הגרף עומד בדרישה שהוא גרף מכוון, ולבסוף אנחנו בודקים האם הגרף לא קיים ברשימת הגרפים שכבר שמרנו (מוודאים שלא שמרנו כבר גרף איזומורפי), ולכן סיבוכיות הזמן על כל גרף יהיה $O(n!)$.

ולכן סיבוכיות הזמן של התוכנית עבור n כלשהוא הינו $O(2^{n(n-1)} \cdot n!)$.

לאחר שעברנו על כל הגרפים האפשריים בגודל n אנחנו שומרים את הרשימה של הגרפים התקינים בקובץ טקסט ושומרים את אורך הזמן שלקח לבצע את כל התהליך

קוד שאלה 2:

בדומה לקוד בשאלה הראשונה אנחנו משתמשים ב-networkx ו-itertools,

בהינתן קובץ טקסט המייצג קשתות בגרף, ראשית אנחנו מפרשים את הקובץ טקסט ויוצרים גרף מתאים לקלט, לאחר מכן אנחנו יוצרים רשימה של כל הגרפים האפשריים בגודל n (גרפים מחוברים לא איזומורפיים).

בשביל לספור את כל המוטיפים הקיימים בגרף אנחנו עוברים על כל הקומבינציות האפשריות של nodes n, אנחנו יוצרים תת-גרף מכל קומבינציה, ואם התת-גרף מחובר אז אנחנו מחפשים לאיזה מוטיפ הוא מתאים ומוסיפים +1 לספירה של כמות הפעמים שהמוטיפ הופיע.

בסוף ההרצה אנחנו מייצאים לקובץ טקסט רשימה של כל המוטיפים האפשריים וכמה פעמים כל מוטיפ הופיע בגרף הנתון

1.ב.

Timing results for motif generation

n=1: 0 motifs found in 0.00 seconds

n=2: 2 motifs found in 0.00 seconds

n=3: 13 motifs found in 0.01 seconds

n=4: 199 motifs found in 4.45 seconds

1.ג. ניסינו להריץ עבור n=5, אבל לאחר שעה הקוד עדיין לא סיים לרוץ, ולכן עבור n=4 הינו המספר המקסימלי עבורו זמן הרציה הינו מקסימום שעה אחת

1.ד. זמן סיבוכיות הקוד הינו: $O(2^{n(n-1)} \cdot n!)$, בעזרת נתון זה וזמני הריצה שקיבלנו נשערך מה יהיה זמן הריצה של n=5

עבור n=4 מספר הגרפים האפשריים הינו $2^{4 \cdot 3}$

עבור n=5 מספר הגרפים האפשריים הינו $2^{5 \cdot 4}$

בנוסף יש לנו את הסיבוכיות האיזומורפית: $\frac{5!}{4!} = 5$

לכן נצפה כי זמן הריצה של n=5 יהיה בסדר גודל של פי $5 = \frac{2^{5 \cdot 4}}{2^{4 \cdot 3}}$ לעומת זמן הריצה של n=4

אזי נשערך כי זמן הריצה עבור n=5 יהיה בסדר גודל של:

$$1280 \cdot 4.45 = 5696[sec] = \frac{5696}{3600}[hours] = 1.6[hours]$$

אזי אנו יודעים כי עבור n=5 הקוד יסיים את זמן ריצתו לאחר סדר גודל של שעות בודדות

ולכן ניתן להסיק, לפי הסיבוכיות שעבור n=6 יקח הרבה יותר מ 8 שעות

או בקיצור עבור n=5 הקוד יסיים את ריצתו בתחום של שעה ל 8 שעות אך עבור n>5 יידרש זמן ריצה שהינו הרבה יותר גדול מ 8 שעות.