רועי אסא 322542226
ייטב נסים 322713488
קישור ל-README.md: https://github.com/itavn/biological-computing-ex1

# שאלה 1 - אלגוריתם למציאת כל תתי גרפים קשירים מגודל n

### 1. יצירת כל הקשתות האפשריות עבור $n$ מסוים

צור סט של כל הקשתות האפשריות בגרף עם $n$ קודקודים (סה"כ מקבלים רשימה בגודל $n^2-n$).
**הסבר:** יש סה"כ $n^2-n$ קשתות מכוונות אפשריות כיוון שניתן לעשות לולאה כפולה על אינדקסי הקודקודים מ-$0,\dots,n-1$ וכל פעם לא מאפשרים קשת עצמית.
הפלט של שלב זה:

$$edges_{list} = [e_1, e_2, \dots e_{n^2-n}] \quad , \quad s.t: \ e_i = [v_k, v_j]$$

### 2. מציאת כל הקומבינציות

מציאת קומבינציות (של קשתות מרשימת הקשתות) מגודל $size \in \{n-1, n, \dots, n^2-n\}$.
עבור כל גודל, נקבל רשימה של קומבינציות אפשריות מהגודל הזה.
כדי לבצע זאת, יש לרוץ בלולאה על כל גדלי הקומבינציות האפשריות $size \in \{n-1, n, \dots, n^2-n\}$ ועבור כל אחד להשתמש בפונקציה itertools.combinations(edges_list, comb_size) כדי לקבל רשימה של כל הקומבינציות האפשריות מגודל $size$ מסוים.
הפלט המתקבל לאחר שלב זה הינו כדלקמן:

$$Combinations_{list} \triangleq [c_{n-1}, c_n, \dots c_{n^2-n}] \quad , \quad s.t: \ c_i = \{All\ combinations\ of\ size\ i\}$$

$$c_i = \left[c_{i_1}, \dots, c_{i_{b_i}}\right] \quad s.t: \ c_{i_k} = \{k^{th}\ combination\ of\ size\ i\} =$$

$$= \{list\ of\ i\ edges\ from\ edges_{list}\} \ \&\& \ b_i \triangleq \binom{n^2-n}{i}$$

כאשר: $b_i$ = מספר הקומבינציות האפשריות עבור קומבינציה מגודל $i$.

### 3. בדיקת קשירות

בדוק האם כל קומבינציה (=תת גרף אופציונלי, דהיינו $c_{i_k}$) מכל גודל $i \in \{n-1, n, \dots, n^2-n\}$ מהווה תת גרף קשיר. כלומר, עלינו לבדוק האם סט הקשתות שנבחרו בכל קומבינציה $c_{i_k}$ בנפרד אכן פורשות את כל $n$ הצמתים בגרף.
נציין כי כל תת גרף קשיר חייב להכיל לפחות $n-1$ קשתות. אבל, זה תנאי הכרחי ולא מספיק. ייתכן שהקשתות שנבחרו לא יוצרים תת גרף קשיר. כמו כן, מספר מקסימלי של קשתות מכוונות בתת גרף הינו $n^2-n$.

### 4. בדיקת איזומורפיזם

עבור כל קומבינציה $c_{i_m}$ (המהווה תת גרף קשיר לאחר 3), נבדוק האם קיימת קומבינציה $c_{i_k}$ עבור $m \neq k$ איזומורפית ל-$c_{i_m}$ כלומר:
$c_{i_m} \approx c_{i_k}$. נבצע את זה ע"י שימוש בפונקציית עזר *is_isomorphic(G1, G2)* מספריית *networkx*.
נעשה זאת ע"י שימוש בלולאה באופן הבא:
⇐ לולאה חיצונית רצה על כל סט של קומבינציות מגודל מסוים $c_{n-1}, \dots, c_{n^2-n}$ מתוך $Combinations_{list}$.
⇐ לולאה פנימית תהיה לולאה מקוננת אשר רצה על כל קומבינציה מתוך סט הקומבינציות מגודל מסוים $c_{i_m}$, ובודקת על כל הקומבינציות האחרות $c_{i_k}$ אם הם איזומורפיות ל-$c_{i_m}$ (עבור $k \neq m$).
**הערה חשובה:** גרפים מקומבינציות מגדלים שונים (לדוגמא $c_{i_k}, c_{m_z}$ עבור $i \neq m$ ) לא יכולים להיות איזומורפיים כיוון שכאשר מספר הקשתות שונה בין צמד גרפים <= הגרפים אינם איזומורפיים.

### 5. איחוד כל הקומבינציות מכל גודל

נותרנו עם רשימה של $Combinations_{list}$ המכילה רשימות שכל אחת מהן מכילה את אוסף הקומבינציות מגודל מסוים שהן לא איזומורפיות והן מהוות גרף קשיר עבור $n$ צמתים.

רועי אסא 322542226
ייטב נסים 322713488
קישור ל-README.md: https://github.com/itavn/biological-computing-ex1

כעת, יש לאחד את כל תתי הרשימות לרשימה אחת שתכיל את כל תתי הגרפים האפשריים מכל הגדלים האפשריים.
מוקדם יותר באלגוריתם יצרנו את הרשימה הבאה:

$$Combinations_{list} = [c_{n-1}, c_n, \dots c_{n^2-n}] = \left[ \underbrace{\left[c_{n-1_1}, \dots, c_{n-1_{b_{n-1}}}\right]}_{\substack{all\ combinations \\ of\ size\ n-1}}, \dots, \underbrace{\left[c_{n^2-n_1}, \dots, c_{n^2-n_{b_{n^2-n}}}\right]}_{\substack{all\ combinations \\ of\ size\ n^2-n}} \right]$$

$$\Downarrow$$

$$All - size - combinations_{list} = \{list\ of\ all\ combinations\ of\ different\ sizes\} = \bigcup_{i=n-1}^{n^2-n} c_i$$

$$\Downarrow$$

$$\boldsymbol{All - size - combinations_{list}} = \bigcup_{i=n-1}^{n^2-n} \left( \underbrace{\left[c_{i_1}, \dots, c_{i_{b_i}}\right]}_{all\ combinations\ of\ size\ i} \right)$$

### 6. *יצירת קובץ* sub_graphs_with_n_nodes.txt

מתוך רשימת כל הקומבינציות מכל הגדלים, דהיינו $\boldsymbol{All - size - combinations_{list}}$, יש ליצור קובץ שבו בהתחלה מדפיסים את $n$, לאחר מכן את $count$ שמונה את גודל הרשימה $\boldsymbol{All - size - combinations_{list}}$ ולבסוף לסרוק כל קומבינציה ולרשום בכל שורה בנפרד את הקשת. אם לדוגמא הקומבינציה מגודל 4, אזי יודפסו 4 הקשתות בקומבינציה, בכל שורה בנפרד בהתאם לדרישה במטלה. עלינו לסרוק כל תת גרף מכל גודל $i \in \{n-1, \dots, n^2-n\}$ ולכתוב בקובץ החדש את הקשתות בתת הגרף בהתאם לדרישת המטלה.

## שאלה 2 – מציאת מספר מופעים של כל $motif$ מגודל $n$ בגרף נתון

בשאלה זו מקבלים מספר שלם $n$ ואת הגרף הנתון בשאלה בצורה של רשימת קשתות.

**הערות:**

- המספר $n$ שמקבלים עשוי להיות שונה ממספר הקודקודים בגרף הקלט, לרוב יהיה נמוך יותר.
- נניח כי את הקלט של הקשתות מקבלים מקובץ חיצוני שנקרא לו $\boldsymbol{edges\_list.txt}$ שממנו נקרא בקוד את הקשתות שורה שורה ובצורה הזו נבנה את הגרף הנתון.
- מניחים כי ייצרנו כבר את הקבצים *sub_graphs_with_n_nodes.txt* כאשר $n \in \{1, 2, 3, 4\}$ לפי הקבצים האלה, נוכל לבנות רשימה של כל תתי הגרפים החוקיים *all_sub_graphs_list*.

**אלגוריתם:**

בהנתן $n$ מסוים, נקרא את הקובץ **sub_graphs_with_n_nodes.*txt*** כדי ליצור רשימה של כל המוטיפים שמצאנו מגודל $n$. עבור גרף נתון עם סט של קשתות, נסמן אותן ב- *edges_list*. לאחר מכן, נמצא את כל הקומבינציות מגודל מסיום של קשתות הנ"ל בגרף (מספר הקשתות שנרצה למצוא גרף איזומורפי אליהן יהיה לפי מספר הקשתות של המוטיף הנוכחי). כזכור, אנחנו רצים על כל המוטיפים, ועבור כל אחד יש לו מספר מסוים של קשתות (לפי השאלה הראשונה). לכן, נרצה למצוא את כל הקומבינציות בגרף הנתון (*test_graph*) עם אותו מספר של קשתות של המוטיף הנוכחי. כפי שהסברנו בשאלה הראשונה, מספר הקומבינציות הנ"ל שווה ל- $c_{motif} = \binom{|\boldsymbol{edges_{list}}|}{|\boldsymbol{edges_{motif}}|}$

כעת, עבור כל קומבינציה אפשרית שקיבלנו, נבנה את הגרף עבורה ונבדוק האם הגרף הנ"ל איזומורפי למוטיף מסויים. ניתן לעשות זאת ע"י הפקודה $nx.is\_isomorphic(G1, G2)$ מתוך ספריית $networkx$.