עבודת בית 4 – תכנון אלגוריתמים 2023

תאריך הגשה: 8/1 את העבודה יש להגיש במודל (עדיף מוקלד)

*מומלץ ביותר לא להמתין לרגע האחרון להגשת העבודה.

מתרגל אחראי: אורי פרידמן

הוראות כלליות:

- כל עוד לא נאמר אחרת, כאשר הנכם מתבקשים לתאר אלגוריתם יש לספק את הבאות:
  1. תיאור מילולי של האלגוריתם.
  2. הוכחת נכונות.
  3. ניתוח זמן ריצה.
- אלגוריתם עם זמן ריצה אקספוננציאלי לא נחשב יעיל ולכן בדרך כלל לא יתקבל.
- פתרון יש לכתוב רק בדף התשובות הנלווה לעבודה.

בהצלחה!

שאלה 1

אלגוריתם דיניץ בקיבולות נמוכים:

בתרגיל אנחנו נראה שאם אנו מגבילים את הקיבול על הקשתות ל1, כלומר $c(u,v)=\begin{cases}1, & (u,v)\in E\\ 0, & else\end{cases}$ אז זמן הריצה של האלגוריתם של דיניץ הוא $O(|E|\sqrt{|E|})$. כשאנו נגיד שלב אנו נתכוון לאיטרציה בודדת של הלולאה באלגוריתם הראשי. אנו נניח $|V|\le|E|$.

סעיף א':

הראו כי עלות של כל שלב עם ההנחה על פונקציית הקיבול היא $O(|E|)$ (ולא $O(|E|\cdot|V|)$ כמו בניתוח המקורי). כלומר הראו שברשת השכבות עבור רשת עם קיבולות $\{0,1\}$ מציאת זרימה חוסמת לוקחת $O(|V|+|E|)$.

בסעיפים הבאים נראה כי מספר השלבים שמתבצע על ידי האלגוריתם הוא לכל היותר $2\sqrt{2|E|}$.

סעיף ב':

תהי $N=((V,E),c,s,t)$ רשת זרימה שבה הקיבולות הם $\{0,1\}$ וf זרימה חוקית ברשת, הוכיחו כי אם המרחק של s מt **ברשת השיורית** הוא ℓ, אזי גודל זרימת מקסימום בN_f הוא לכל היותר $\frac{2|E|}{\ell}$.

הדרכה: הראו דרך רשת השכבות שקיים חתך בN_f שקיבולו קטן או שווה ל$\frac{|E_f|}{\ell}$.

סעיף ג':

הסבירו מדוע לכל ℓ, אחרי לכל היותר ℓ שלבים המרחק בין s ל-t הוא לפחות ℓ. (זה סעיף מכווין.)

סעיף ד':

תהי f זרימה ברשת זרימה ברשת N. הוכיחו כי אם בשלב הi גודל זרימת המקסימום בN_f קטן או שווה $\frac{2|E|}{\ell}$, אזי אחרי לכל היותר $i+\frac{2|E|}{\ell}$ שלבים האלגוריתם של דיניץ מוצא זרימת מקסימום.

ניתן להשתמש בטענה הבאה ללא הוכחה:

טענה: אם f זרימה ברשת N ו-g זרימת מקסימום ברשת השיורית N_f אזי $f+g$ היא זרימת מקסימום ברשת N וגודלה

$|f|+|g|$.

סעיף ה':

הראו כי מספר השלבים שנעשה בכל ריצת האלגוריתם על רשת עם קיבולות $\{0,1\}$ הוא לכל היותר $2\sqrt{2|E|}$.

הדרכה: בחרו ℓ מתאים עבור סעיפים ב',ד'.

סעיף ו':

הסבירו מדוע זמן הריצה של אלגוריתם דיניץ עבור קיבולות 0,1 הוא $O(|E|\sqrt{|E|})$.

שאלה 2

יש לנו סיכוי?

בליגת כדורסל בעיר כלשהי הקבוצות משחקות אחת מול השנייה מספר רב של פעמים. הקבוצות עם מספר הניצחונות הגדול ביותר מנצחות את הליגה (ייתכן יותר מקבוצה אחת). בכדורסל כידוע אין תוצאות תיקו. אחרי שכל הקבוצות שיחקו במהלך העונה מצב הליגה ניראה כך:

קבוצה	ניצחונות
מכבי אריות	75
הפועל שועלים	71
הקופים הירוקים	69
אליצור גמלים	63
מכבי נמיות	?

בנוסף נתונה הטבלה של כמה מפגשים נשארו לכל הקבוצות לחשק אחת מול השנייה.

	כמה משחקים נשאר לכל קבוצה לשחק אחת מול השנייה				
מכבי אריות	-	5	7	4	7
הפועל שועלים	5	-	2	4	7
הקופים הירוקים	7	2	-	4	7
אליצור גמלים	4	4	4	-	7
מכבי נמיות	7	7	7	7	-
	מכבי אריות	הפועל שועלים	הקופים הירוקים	אליצור גמלים	מכבי נמיות

אם למכבי נמיות יש 46 ניצחונות אז ברור שאין להם סיכוי לזכות בליגה כי אפילו אם הם ינצחו בכל המשחקים יהיה להם 74 ניצחונות ולמכבי אריות יש 75 ניצחונות. מה קורה אם יש להם 47 ניצחונות? עדיין אין להם סיכוי מכיוון שלמכבי אריות והפועל שועלים נשארו 5 משחקים אחת מול השנייה או שמכבי אריות תנצח לפחות משחק אחד ותגיע ל76 ניצחונות או שהפועל שועלים תנצח את כל ה5 ותגיע ל76 ניצחונות. כלומר, אחת מהן תגיע ל76 ניצחונות, אפילו אם מכבי נמיות ינצחו בכל המשחקים יהיו להם 75 ניצחונות. אבל מה קורה כאשר יש לנמיות $w_5 = 48$ ניצחונות?

נראה איך לפתור שאלה זו ע"י שימוש בזרימת מקסימום.

נגדיר את הבעיה באופן פורמאלי:

קלט: מערך T של מספר הניצחונות של הקבוצות בליגה ממוין בסדר יורד, כלומר, $T[1] \geq T[2] \geq \cdots \geq T[n]$ ומטריצה סימטרית של המשחקים הנותרים R שעבור $1 \leq i,j \leq n$ התא $R[i,j]$ שווה לכמות המשחקים שנשארו לקבוצות i,j לשחק בעתיד אחת מול השנייה, כאשר $R[i,i] = 0$ ומתקיים $R[i,j] = R[j,i]$.

נסמן $r_i = \sum_{j=1}^{n} R[i,j]$, מספר המשחקים שנותרו לקבוצה הi.

פתרון חוקי: מטריצה C שהיא חלוקה של הניצחונות הנותרים בR, כלומר הכניסה $C[i,j]$ היא מספר הניצחונות של הקבוצה הi על הקבוצה הj כאשר לכל i,j מתקיים $C[i,j] + C[j,i] = R[i,j]$, (כלומר, בכל משחק בין i לj יש מנצחת יחידה) ו $C[i,j] \geq 0$.

נסמן $S[i] = T[i] + \sum_{j=1}^{n} C[i,j]$, את מספר המשחקים שהקבוצה תנצח בסוף העונה.

יש למצוא: האם קיים פתרון חוקי C כך ש $S[n] \geq S[j]$ לכל $1 \leq j \leq n-1$, כלומר שבסוף העונה הקבוצה האחרונה תהייה במקום הראשון.

כדי לפתור את הבעיה נגדיר רשת זרימה באופן הבא:

הקודקודים הם: $V = \{s,t\} \cup \{b_1, \dots, b_{n-1}\} \cup \{a_{m,\ell} \mid 1 \leq m < \ell \leq n-1\}$.

הקשתות הן:

$$E_1 = \{(s, a_{m,\ell}) \mid 1 \leq m < \ell \leq n-1\}$$

$$E_2 = \{(a_{m,\ell}, b_m), (a_{m,\ell}, b_\ell)\} \mid 1 \leq m < \ell \leq n-1\}$$

$$E_3 = \{(b_m, t) \mid 1 \leq m \leq n-1\}$$

$$.E = E_1 \cup E_2 \cup E_3$$

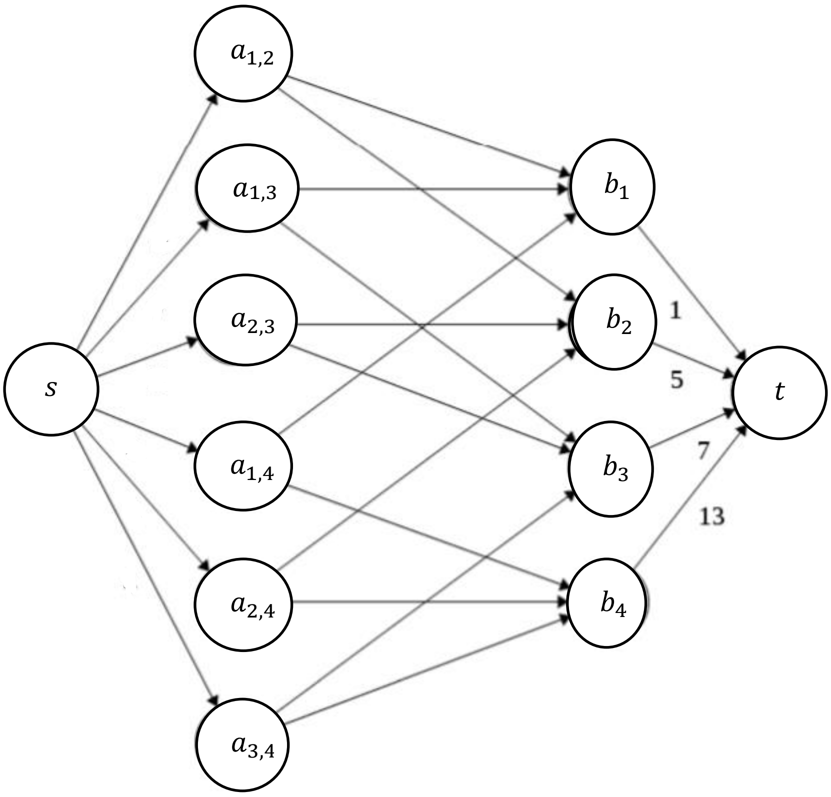
אנו נגדיר את פונקציית הקיבול על הקשתות השייכות לE_3 בדרך הבאה

$$c(b_m, t) = T[n] + r_n - T[m]$$

(כלומר, מספר הניצחונות המקסימאלי של הקבוצה הn פחות כמות הניצחונות הנוכחית של הקבוצה הm). בשלב הזה אם נגלה כי הקיבול של אחת הקשתות קטן מאפס אז הקבוצה הn לא תוכל לנצח.

את שאר הקיבולות תגדירו בהמשך.

קודקוד המקור וקודקוד הבור הם s,t והרשת תהיה $N = ((V,E), c, s, t)$.
בעמוד הבא מתוארת הרשת עבור המקרה בדוגמה.



לדוגמה הקיבול של הקשת (b_3, t) היא $T[5] + r_5 - T[3] = 48 + 28 - 69 = 7$ ומייצג כמה משחקים תוכל לנצח קבוצה 3 ואנו עדיין נוכל לנצח לפחות כמותה.

סעיף א':

הגדירו את הקיבולות על הקשתות מE_1, E_2 כך שהקבוצה הn תוכל לזכות בליגה אם ורק אם קיימת זרימת מקסימום המרווה את כל הקשתות היוצאות מs כלומר $c\left(s, a_{m,\ell}\right) = f\left(s, a_{m,\ell}\right)$ לכל $1 \leq m < \ell \leq n - 1$. הניחו כי הקבוצה ה$n$ מנצחת את כל המשחקים שנשארו לה, והוכיחו כי עם הקיבולים האלו הקבוצה הn יכולה לזכות בליגה אם ורק אם קיימת זרימת מקסימום המרווה את כל הקשתות היוצאות מs.

סעיף ב':

בעזרת סעיף א' קיבעו אם למכבי נמיות יש סיכוי לזכות בליגה, כאשר יש לה 48 ניצחונות.

הדרכה: ניתן לעשות זאת בעזרת מציאת זרימה המרווה את הקשתות היוצאות מs או בעזרת חתך שהגודל שלו קטן ממש מסכום הקשתות היוצא מs.

שאלה 3

תת מחרוזת משותפת ארוכה ביותר

בשאלה זאת נפתור את בעיית אורך המחרוזות המשותפת הארוכה ביותר בעזרת אלגוריתם אקראי.

תזכורת:

הגדרה $z = (z_1, \dots, z_k) \in \{0,1\}^k$ היא תת מחרוזת של $x = (x_1, \dots, x_n) \in \{0,1\}^n$ אם קיים אינדקס i כך ש $x_i = z_1, x_{i+1} = z_2, \dots, x_{i+k-1} = z_k$ כלומר z מופיע בx בצורה רציפה.

בעיית תת מחרוזת משותפת ארוכה ביותר:

<u>**קלט:**</u> שתי מחרוזות $x = (x_1, \dots, x_n) \in \{0,1\}^n$ ו– $y = (y_1, \dots, y_m) \in \{0,1\}^m$, כאשר $m < n$.

<u>**פתרון חוקי:**</u> מחרוזת $z = (z_1, \dots, z_k)$ שהיא תת מחרוזת של $x = (x_1, \dots, x_n)$ ושל $y = (y_1, \dots, y_m)$.
מחרוזת זאת נקראת תת מחרוזת משותפת.

<u>**יש למצוא:**</u> **אורך** מחרוזת משותפת ארוכה ביותר.

אתם בטח יודעים שעל ידי תכנון דינאמי ניתן לפתור את הבעיה בעזרת אלגוריתם דטרמיניסטי הרץ בזמן $O(n \cdot m)$.

אנו נפתור את הבעיה בעזרת אלגוריתם הסתברותי יעיל יותר.

<u>**סעיף א':**</u>

נגדיר את בעיית תת המחרוזת המשותפת באורך ℓ.

<u>**קלט:**</u> שתי מחרוזות $x = (x_1, \dots, x_n) \in \{0,1\}^n$ ו$y = (y_1, \dots, y_m) \in \{0,1\}^m$ ומספר טבעי $\ell > 0$.

<u>**פלט:**</u> האם קיימת מחרוזת משותפת באורך ℓ.

תארו אלגוריתם **אקראי** המקיים:

1) אם קיימת מחרוזת משותפת באורך ℓ אזי האלגוריתם מחזיר תמיד כן.
2) אם לא קיימת מחרוזת משותפת באורך ℓ אזי האלגוריתם טועה, כלומר מחזיר כן, בהסתברות של לכל היותר $\frac{1}{4\log(n)}$.
3) זמן הריצה של האלגוריתם הוא $O\big(n \cdot \log(n)\big)$.

הוכיחו נכונות והסבירו כיצד לממש את האלגוריתם בזמן הנדרש.

הדרכה: השתמשו ברעיונות מהאלגוריתם להתאמת מחרוזות שראיתם בהרצאה, בחרו q מספיק גדול על מנת לקבל את השגיאה הנדרשת.

<u>**סעיף ב':**</u>

תארו אלגוריתם הסתברותי שבהסתברות גדולה או שווה מ$\frac{3}{4}$ מוצא את אורך מחרוזת משותפת ארוכה ביותר בזמן של $O(n \cdot \log^2(n))$. ונתחו את זמן הריצה שלו.

רמז: השתמשו בחיפוש בינארי.

שאלה 4

בעיית מסלול פשוט באורך k:

בבעיית k-path אנו רוצים למצוא אם קיים בגרף מכוון G מסלול פשוט $P =< v_1, \dots, v_k >$ כלומר מסלול העובר בk קודקודים שונים זה מזה.

תיאור הבעיה:

<u>קלט:</u> גרף מכוון $G = (V, E)$ ומספר $k \leq \ln(n)$, כש-n הוא מספר הקודקודים בגרף.

<u>פתרון:</u> יש למצוא אם קיים מסלול פשוט $P =< v_1, \dots, v_k >$ בG כלומר $v_i \neq v_j$ לכל $1 \leq i < j \leq k$.

<u>סעיף א':</u>

תארו אלגוריתם **אקראי** עם שגיאה חד כיוונית לבעיה, כאשר אם אין מסלול פשוט עם k קודקודים אז האלגוריתם תמיד יחזיר "אין מסלול" ואם קיים מסלול הוא מחזיר "קיים" בהסתברות של לפחות $\frac{k!}{k^k}$.

הדרכה: השתמשו <u>כקופסה שחורה</u> באלגוריתם למציאת מסלול k צבעוני ממטלה 2.

<u>סעיף ב':</u>

כמה פעמים צריך לחזור על האלגוריתם מסעיף א' באופן בלתי תלוי כך שאם קיים מסלול פשוט עם k קודקודים אז בהסברות של לפחות 0.99 באחת החזרות של האלגוריתם אנחנו נחזיר קיים.

השתמשו בעובדות: $k! > \left(\frac{k}{e}\right)^k$ ו-$\left(1 - \frac{1}{n}\right)^{bn} \leq \frac{1}{e^b}$ עבור $b \geq 1$, אין צורך להוכיח עובדות אלו.

<u>סעיף ג':</u>

השתמשו בסעיפים א' וב' כדי לתאר אלגוריתם הסתברותי עם שגיאה חד כיוונית כך שאם קיים מסלול פשוט עם k קודקודים אז הוא מחזיר כן בהסברות של לפחות 0.99 ואם לא קיים מסלול כנ"ל הוא תמיד מחזיר שאין מסלול. על האלגוריתם לרוץ בזמן $O\left(|V|^2(|E| + |V|)\right)$. הוכיחו נכונות האלגוריתם ונתחו את זמן ריצתו.