# הרצאה 22: חסמי צ'רנוף ומבוא לאלגוריתמי קירוב

*Textbook: Cormen, Leiserson, Rivest, Stein. Introduction to Algorithms.*

מרצה: אורי שטמר

**<u>תזכורת - חסמי צ'רנוף והופדינג:</u>**

יהיו $X_1, X_2, \dots, X_n$ משתנים מקריים בלתי תלויים כאשר לכל $i$ מתקיים $\Pr[X_i = 1] = p$ ו- $\Pr[X_i = 0] = 1 - p$, עבור פרמטר $0 < p < 1$. תוחלת סכום המשתנים היא:

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = p \cdot n$$

(א) לכל $0 < \delta < 1$ מתקיים $\Pr[\sum_{i=1}^{n} X_i \geq (1+\delta)\cdot pn] < \exp(-\delta^2 pn/4)$

(ב) לכל $0 < \delta < 1$ מתקיים $\Pr[\sum_{i=1}^{n} X_i \leq (1-\delta)\cdot pn] < \exp(-\delta^2 pn/2)$

(ג) לכל $\delta > 0$ מתקיים $\Pr[|\sum_{i=1}^{n} X_i - pn| \geq \delta] \leq 2\exp(-2\delta^2/n)$

**<u>עוד סיפור מוטיבציה:</u>**

נניח שאנחנו מטילים מטבע 1000 פעמים. למרות <u>שתוחלת</u> מספר הפעמים שנקבל heads היא 500, <u>ההסתברות</u> שבאמת נקבל 500 (בדיוק) היא מאוד קטנה. מה כן אפשר להגיד?

- אנחנו יכולים להיות "דיי בטוחים" שזה יהיה בין 450 ל- 550 (ההסתברות זה לא קורה היא $\leq 0.02$)
- אנחנו יכולים להיות אפילו "יותר בטוחים" שזה יהיה בין 440 ל- 560 (ההסתברות זה לא קורה היא לכל היותר 0.0015)

איך אפשר להראות את זה בעזרת חסמי צ'רנוף/הופדינג?

נגדיר משתנים מקריים $X_1, \dots, X_n$ עבור $n = 1000$ כאשר לכל $i$:

$X_i = 1$ אם ורק אם בהטלה ה- $i$ יצא heads (ואחרת $X_i = 0$)

נשים לב שלכל $i$ מתקיים $\Pr[X_i = 1] = \Pr[X_1 = 0] = 1/2$ ובנוסף $X_1, \dots, X_n$ הם בלתי תלויים מכיוון שהטלות המטבע נעשות באופן בלתי תלוי. בעזרת חסם הופדינג נקבל:

$$\Pr\left[\begin{matrix}\text{מספר ה } heads \\ \text{לא בין 450 ל 550}\end{matrix}\right] = \Pr\left[\left|\sum_{i=1}^{n} X_i - 500\right| \geq 50\right] \leq 2\cdot\exp\left(-\frac{2\cdot 50^2}{1000}\right) = 2\cdot\exp(-5) \leq 0.014$$

## **הוכחת חסמי צ'רנוף/הופדינג**

אנחנו נוכיח אק את (א). ההוכחות של (ב) ושל (ג) מאוד דומות. לפני שנתחיל בהוכחה, נזכר ב- 2 דברים מהסתברות:

**תזכורת 1 – אי איוויון מרקוב:**
**לכל משתנה מקרי אי שלילי $Y$ ולכל $0 < a$ מתקיים $\Pr[Y \geq a] \leq \frac{1}{a} \cdot \mathbb{E}[Y]$**

הוכחה עבור משתנה מקרי $Y$ בדיד:

$$\mathbb{E}[Y] = \sum_{y} y \cdot \Pr[Y = y] \geq \sum_{y \geq a} y \cdot \Pr[Y = y] \geq \sum_{y \geq a} a \cdot \Pr[Y = y] = a \cdot \sum_{y \geq a} \Pr[Y = y] = a \cdot \Pr[Y \geq a]$$

**תזכורת 2 – עבור משתנים מקריים <u>בלתי תלויים</u> מתקיים שתוחלת של מכפלה שווה לתוחלת של מכפלות, כלומר $\mathbb{E}[X_1 \cdot X_2 \cdots X_n] = \mathbb{E}[X_1] \cdot \mathbb{E}[X_2] \cdots \mathbb{E}[X_n]$**

הוכחה עבור משתנים מקריים $X, Y$ בדידים ובלתי תלויים:

$$\mathbb{E}[X \cdot Y] = \sum_{x.y} \Pr[X = x, Y = y] \cdot xy = \sum_{x.y} \Pr[X = x] \cdot \Pr[Y = y] \cdot xy$$

$$= \left(\sum_{x} \Pr[X = x] \cdot x\right) \cdot \left(\sum_{y} \Pr[Y = y] \cdot y\right) = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

## **נחזור להוכחת גרסה (א) של חסם צ'רנוף:**

יהיו $X_1, X_2, \ldots, X_n$ משתנים מקריים בלתי תלויים כאשר לכל $i$ מתקיים $\Pr[X_i = 1] = p$ ו- $\Pr[X_i = 0] = 1 - p$
ויהי $0 < \delta < 1$. עלינו להראות ש-

$$\Pr\left[\sum_{i=1}^{n} X_i \geq (1 + \delta) \cdot pn\right] < \exp(-\delta^2 pn/4)$$

נסמן $t = (1 + \delta)pn$ ונסמן $c = \delta/2$.
(נשים לב שמכיוון ש- $0 < \delta < 1$ אזי $0 < c < 1/2$)

נחשב:

$$\Pr[\sum X_i \geq t] = \Pr[c \cdot \sum X_i \geq c \cdot t] = \Pr\left[e^{c \cdot \sum X_i} \geq e^{c \cdot t}\right] = ((1))$$

כעת לפי אי שוויון מרקוב נקבל ש

$$\big((1)\big) \leq e^{-c \cdot t} \cdot \mathbb{E}\left[e^{c \cdot \sum X_i}\right] = e^{-c \cdot t} \cdot \mathbb{E}[e^{c \cdot X_1} \cdot e^{c \cdot X_2} \cdots e^{c \cdot X_n}] = ((2))$$

כעת מכיוון שהמשתנים $X_1, \ldots, X_n$ הם בלתי תלויים נקבל ש

$$\big((2)\big) = e^{-c \cdot t} \cdot \mathbb{E}[e^{c \cdot X_1}] \cdot \mathbb{E}[e^{c \cdot X_2}] \cdots \cdot \mathbb{E}[e^{c \cdot X_n}] = ((3))$$

ומכיוון ש- $X_1, \ldots, X_n$ מתפלגים אותו הדבר מתקיים $\mathbb{E}[e^{c \cdot X_1}] = \mathbb{E}[e^{c \cdot X_2}] = \cdots = \mathbb{E}[e^{c \cdot X_n}]$ ולכן

$$\big((3)\big) = e^{-c\cdot t} \cdot \left(\mathbb{E}[e^{c\cdot X_1}]\right)^n$$

כלומר קיבלנו ש

$$\Pr[\sum X_i \geq t] \leq e^{-c\cdot t} \cdot \left(\mathbb{E}[e^{c\cdot X_1}]\right)^n$$

כעת נשים לב ש-

$$\mathbb{E}[e^{c\cdot X_1}] = p \cdot e^c + (1-p)e^0 = p \cdot e^c + 1 - p \leq p(1 + c + c^2) + 1 - p = 1 + p(c + c^2) \leq e^{p(c+c^2)}$$

כאשר אי השיוויון הראשון הוא לפי הנוסחה $e^z \leq 1 + z + z^2$ המתקיימת לכל $z \leq 1$ וכאשר אי השיוויון השני הוא לפי הנוסחה $1 + z \leq e^z$ המתקיימת לכל $z \in \mathbb{R}$.

נציב זאת בחזבון הקודם שעשינו ונקבל

$$\Pr[\sum X_i \geq t] \leq e^{-c\cdot t} \cdot \left(\mathbb{E}[e^{c\cdot X_1}]\right)^n \leq e^{-c\cdot t} \cdot \left(e^{p(c+c^2)}\right)^n = e^{-c\cdot t} \cdot e^{pn(c+c^2)} = \big((4)\big)$$

נזכור שבחרנו $t = (1+\delta)pn$ ולכן נקבל

$$\big((4)\big) = e^{-c\cdot(1+\delta)pn} \cdot e^{pn(c+c^2)} = e^{-cpn(\delta-c)} = \big((5)\big)$$

נזכור שבחרנו $c = \delta/2$ ולכן

$$\big((5)\big) = e^{-\delta^2 pn/4}$$

מ.ש.ל.

## מה קרה פה?

אי-שיוויון מרקוב אומר שמשתנה מקרי אי שלילי לא יכול לסטות בהרבה מהתוחלת שלו. אבל בחסם צ'רנוף רצינו להראות ש- $\sum X_i$ לא יכול לסטות אפילו בקצת מהתוחלת שלו. לכן במקום להפעיל או אי-שיוויון מרקוב ישירות על $\sum X_i$ חשבנו על המשתנה המקרי $\exp(\sum X_i)$.

למה זה טוב? עכשיו מרקוב אומר לנו שההסתברות ש- $\exp(\sum X_i)$ יחרוג בהרבה מהתוחלת שלו היא קטנה וזאת אומרת שההסתברות ש- $\sum X_i$ יחרוג מהתוחלת שלו אפילו בקצת היא קטנה (כי אם $\sum X_i$ חורג אפילו קצת אז $\exp(\sum X_i)$ חורג הרבה...)

בנוסף, בזכות העובדה ש- $X_1, \dots, X_n$ הם בלתי תלויים יכולנו לנתח את התוחלת של $\exp(\sum X_i)$ כי התוחלת הזאת התפצלה לנו למכפלה של תוחלות.

# מבוא לאלגוריתמי קירוב

## הקדמה

עד כה ראינו בקורס בעיות להן קיים אלגוריתם יעיל, אלגוריתם פולינומי שפותר אותן, ולכן ניתן להתייחס לבעיות אלו כאל בעיות "קלות". האם לכל בעיה קיים אלגוריתם יעיל?
ישנן בעיות רבות מאוד להן לא ידוע (ומאמינים כי לא קיים) אלגוריתם פולינומי הפותר אותן. את חלקן אתם תפגשו בקורס "מודלים חישוביים" בסמסטר הבא.

דוגמאות לבעיות אופטימיזציה:

| **בעיה קלה** | **בעיה קשה** |
| --- | --- |
| מציאת מסלול קצר ביותר מ-s ל-t בגרף | מציאת מסלול פשוט ארוך ביותר מ-s ל-t בגרף |
| חתך בעל קיבול מינימום (מוצאים בעזרת אלגוריתם למציאת זרימה מקסימלית) | חתך בעל קיבול מקסימום (גם במקרה הלא ממושקל, עבור חתך בגודל מקסימום עם קיבולים 1) |
| תכנון ליניארי | תכנון בשלמים |
| כיסוי בצמתים בגודל מינימום בגרף דו צדדי (ראינו אלגוריתם מבוסס תכנון ליניארי) | כיסוי בצמתים בגודל מינימום בגרף כללי (נרחיב עליה בשיעורים הבאים) |

## מוטיבציה

למרות שלבעיות קשות לא ידוע אלגוריתם פולינומי הפותר את *המקרה הכללי*, לעיתים אנו נדרשים לפתור מופעים מסויימים של בעיות אלו. מה ניתן לעשות?

<u>אסטרטגיות פתרון אפשריות</u>:

1. להשתמש באלגוריתמים אקספוננציאלים מתוחכמים אשר על קלטים קטנים מספיק רצים בזמן סביר.
   <u>דוגמה</u>: בעית מסלול פשוט ארוך ביותר בגרף.
   - אלגוריתם נאיבי יחשב את כל $n!$ הפרמוטציות האפשריות עבור קודקודי בגרף ויבדוק מי מהן מהווה מסלול ארוך ביותר. שימוש באלגוריתם כזה אפשרי עבור ערכי $n$ קטנים במיוחד.
   - לבעיה זו קיים אלגוריתם אקספוננציאלי (מבוסס תכנון דינאמי בזמן ריצה של פולינום כפול $2^n$). עובדה זו משנה בפועל את גודל הקלטים אשר אפשר לפתור/להתמודד איתם, כיוון ש- $n!$ עולה מהר מאוד ביחס ל- $2^n$ כבר עבור ערכי $n$ קטנים יחסית. בכך הגדלנו את ערכי $n$ המקסימליים איתם אפשר להתמודד.

   <u>דוגמה</u>: (לקוחה מתחום סיבוכיות פרמטרית) מציאת כיסוי בקודקודים בגודל שלא עולה על $k$.
   - אלגוריתם נאיבי יחשב ב- $O(n^k)$ את כל $\binom{n}{k}$ תתי הקבוצות האפשריות ולכל קבוצה יבדוק האם היא כיסוי חוקי.
   - לבעיה זו קיים אלגוריתם הרץ בזמן $2^k(|V|+|E|)$. זהו שיפור גדול ביחס ל- $n^k$. (נביט בעץ החיפוש הבא: כל עוד נותרה קשת $(u,v)$ (לא מכוסה) בגרף, נתפצל לשני מקרים: נסיר את $u$ או את $v$ (ואת כל הקשתות שחלות בהם) ונקטין את $k$ ב-1. נעצור כאשר (א) $k$ שלילי ונחזיר "לא קיים כיסוי" (כי עדיין נותרו קשתות לא מכוסות ע"י הקודקודים שנבחרו שנעשו עד כה) או כאשר (ב) לא נותרו קשתות (לא מכוסות) בגרף.

2. היוריסטיקות.
היוריסטיקה היא דרך פעולה פרקטית שאמורה להביא (אנו מקווים שתביא) לתוצאות טובות מספיק על קלטים אמיתיים מסויימים. במקרה הכללי, ובמיוחד במקרה הגרוע, ייתכן ואף סביר שדרך פעולה זו לא תעבוד כלל. יתרה מכך, לא ניתן להבטיח הבטחות כלשהן (לגבי זמן ריצה או נכונות) עבור קלטים עליהן היא אמורה לעבוד.
דוגמה: בבעיית הפעילויות (לה ראינו אלגוריתם חמדן אשר בחר בכל שלב קטע עם זמן סיום מוקדם ביותר) היוריסטיקה אינטואיטיבית מאוד תהיה לבחור בכל שלב קטע שנחתך עם כמה שפחות קטעים אחרים. גישה זו תעבוד בפועל על דוגמאות רבות, ותניב פתרונות חוקיים גדולים למדיי ואף אופטימאליים במקרים מסויימים, אך בהחלט יכולה להיכשל עבור דוגמאות מסויימות (ראינו אחת בהרצאה).

3. אלגוריתמי קירוב.
ישנן בעיות להן קיים אלגוריתם יעיל המבטיח לנו למצוא פתרון שאינו "גדול מדיי" ("קטן מדיי") ביחס לגודל הפתרון המינימלי (המקסימלי). במילים אחרות, לאלגוריתם זמן ריצה יעיל ומובטח לנו כי יוחזר פתרון "טוב אך לא בהכרח אופטימלי". ייתכן כי יוחזר פתרון אופטימאלי אך לא ניתן להבטיח זאת במקרה הכללי.

בשיעורים הבאים נדבר על אלגוריתמי קירוב. לפני שנצלול לעומק הפרק, נראה דוגמה לאלגוריתם קירוב חמדן ופשוט עבור בעיה אותה פגשנו.

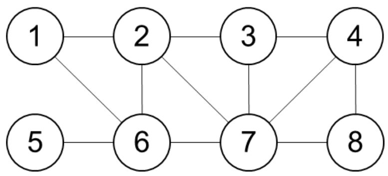
**בעיית כיסוי בקודקודים בגרף כללי לא ממושקל**
מופע: גרף לא מכוון $G = (V, E)$.
פתרון חוקי: תת קבוצה של קודקודים $C \subseteq V$ כך שלכל קשת $(u, v) \in E$ מתקיים כי $u \in C$ או $v \in C$ (או שניהם).
יש למצוא: פתרון חוקי $C$ בגודל מינימום.

הערה חשובה: בפרק על תכנון ליניארי ראינו כי קיים פתרון פולינומי עבור הגרסה הממושקלת **בגרף דו צדדי**.
ראינו גם את המקרה הפרטי בו המשקלים היו 1 ואז משקל הכיסוי ייצג את גודל הכיסוי.
במקרה בו הגרף אינו דו צדדי אלא גרף כללי - לא ידוע אלגוריתם פולינומי. אנו מאמינים כי לא קיים כזה אלגוריתם ויתרה מכך, בעיה זו נחשבת לבעיה קשה שאם יימצא לה אלגוריתם פולינומי נוכל בעזרתו לפתור משפחה שלמה של בעיות קשות שלא ידוע עבורן אלגוריתם יעיל.

שאלה: מדוע האלגוריתם שראינו עבור גרף דו-צדדי לא יעבוד במקרה הכללי?
תשובה: במקרה של גרף כללי התוכנית הליניארית והתוכנית בשלמים לא שקולות.

אלגוריתם קירוב לבעית כיסוי בקודקודים
1. אתחול: $C \leftarrow \emptyset$ (הפתרון הנבנה ע"י האלגוריתם)
2. כל עוד $E \neq \emptyset$ בצע:
   2.1. בחר קשת כלשהי $(u, v) \in E$ ובצע $C \leftarrow C \cup \{u, v\}$
   2.2. הסר את הקודקודים $u, v$ ואת כל הקשתות החלות בהם מן הגרף
3. החזר את $C$

דוגמה:




יהי $G = (V, E)$ הגרף הבא:
$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8\}$

כיסוי מינימלי $C = \{v_2, v_4, v_6, v_7\}$ בגודל 4.

נשקול 3 ריצות אפשריות:
1. נבחר את הקשתות $\{(v_2, v_6), (v_4, v_7)\}$ ונקבל כיסוי בגודל 4.
2. נבחר את הקשתות $\{(v_2, v_6), (v_3, v_7), (v_4, v_8)\}$ ונקבל כיסוי בגודל 6.
3. נבחר את הקשתות $\{(v_1, v_2), (v_3, v_4), (v_5, v_6), (v_7, v_8)\}$ ונקבל כיסוי בגודל 8.

מה שקרה כאן הוא לא מקרי, והפתרונות שיתקבלו לעולם לא יהיו גדולים מ-8.

**<u>משפט (נוכיח בהרצאה הבאה):</u>** אלגוריתם הקירוב הנ"ל תמיד יחזיר פתרון המכיל לא יותר מפי שתיים קודקודים מהמינימום האפשרי.

**<u>הגדרה:</u>**
אלגוריתם לבעיית מקסימיזציה (מינימיזציה) $P$ ייקרא **אלגוריתם קירוב-$\alpha$** אם מתקיימים התנאים הבאים:
1. זמן ריצת האלגוריתם פולינומי בגודל ייצוג הקלט.
2. בהינתן קלט $x$ עבורו קיים פתרון אופטימלי $y$, האלגוריתם יחזיר פתרון חוקי $z$ עם ערך $val(z)$ כך שמתקיים

$$val(z) \geq \alpha \cdot val(y)$$

$$\left( \text{או} \quad val(z) \leq \alpha \cdot val(y) \right)$$

בבעיית מקסימיזציה $\alpha < 1$ ובבעיית מינימיזציה $\alpha > 1$.

דוגמאות:
- בעיית הכיסוי בקודקודים היא בעיית מינימיזציה, בה ציינו כי $\alpha = 2$, כלומר האלגוריתם יחזיר כיסוי שלא עולה בגודלו על פי 2 מהגודל המינימלי האפשרי. בדוגמה לעולם לא יהיה גדול מ- 8.
- אם היינו עוסקים בבעית מקסימום, לדוגמה, מציאת חתך מקסימום (חתך עם מספר גדול ביותר של קשתות חוצות), אז עבור $\alpha = \frac{1}{2}$ לדוגמה, אלגוריתם קירוב-$\alpha$ היה מחזיר חתך שמספר הקשתות החוצות אותו הוא לפחות חצי ממספר הקשתות בחתך מקסימום.

שאלה: מה קורה כאשר $\alpha = 1$?
תשובה: אז מתקבל פתרון בעל ערך אופטימלי.

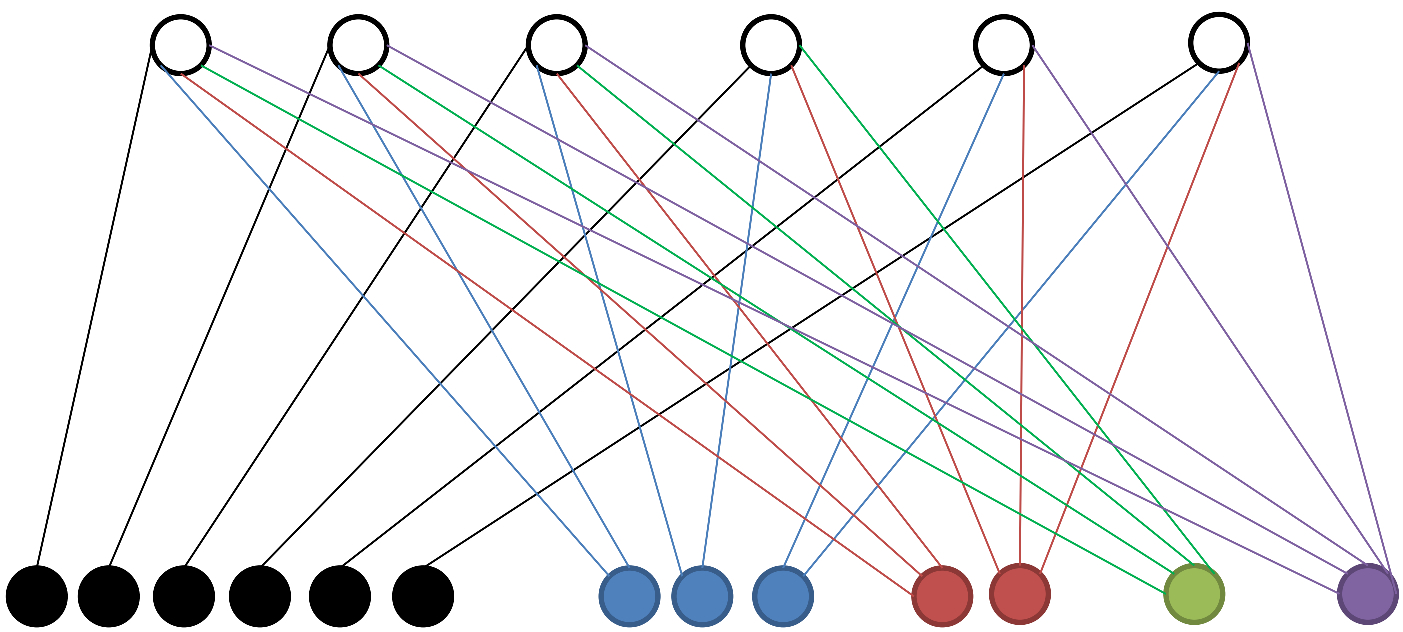
שאלה נוספת: אולי ננסה גם את אלגוריתם הקירוב החמדן הבא לבעית כיסוי בקודקודים:

| |
|---|
| בכל שלב נבחר קודקוד עם דרגה (נוכחית) מקסימלית, נוסיף אותו לכיסויי, ונמחוק מהגרף את כל הקשתות שנוגעות בו. |

האם יחס גם עבור האלגוריתם הזה נקבל יחס קירוב 2 ?

תשובה: לא. הנה דוגמה נגדית.

נבנה גרף דו-צדדי באופן הבא:



הסבר: למעלה יש 6 קודקודים ולמטה יש 13 קודקודים.

נאתחל i=1.
נעבור על הקודקודים התחתונים משמאל לימין בצורה ציקלית.
נחבר את הקודקוד הנוכחי ל- i קודקודים מלמעלה (לפי הסדר).
בכל פעם שמשלימים סיבוב ביחס לקודקודים העליונים אזי נגדיל את i ב 1.

כיסויי אופטימלי הוא כמובן בגודל 6, אבל האלגוריתם החמדן עשויי לבחור בכל פעם את הקודקוד התחתון משמאל וכך לבחור בסופו של דבר את כל התחתונים, כלומר לקבל כיסוי בגודל 13.