תכנון וניתוח אלגוריתמים – הרצאה מס' 7

מרצה: פרופ' בועז פת-שמיר.

משכתב: איליה ריבקין.

בעיית כיסוי קבוצות עם מחירים:

קלט: אוסף $C=\{S_1,\dots,S_n\}$ של קבוצות, מחיר $\cos t(S_i)$ לכל קבוצה i.

פלט: כיסוי Z של $\bigcup_{i=1}^{m} S_i$ ע"י קבוצות מ-C (כאשר $|Z|=n$ הוא סך כל האיברים בקבוצת האיחוד).

המטרה: למזער את עלות הכיסוי הכוללת $\sum_{S\in Z}\cos t(S)$.

פתרון: אלגוריתם חמדן לפי משקל סגולי:

$$X \longleftarrow \bigcup_{i=1}^{m} S_i$$
$$While(X \neq \phi)$$
$$\quad S \longleftarrow \arg\min\left\{\frac{\cos t(S)}{|X\cap S|} : S\cap X \neq \phi\right\}$$
$$\quad \text{Add S to coverage}$$
$$\quad X \longleftarrow X - S$$
$$end$$

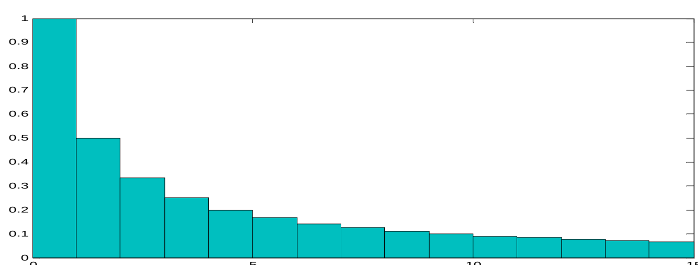
כפי שנראה בהמשך, מקדם הקירוב של האלגוריתם החמדן הוא לוגריתמי.

תחילה נסתכל על הטור ההרמוני הבא:

$$H_n=\sum_{i=1}^{n}\frac{1}{i}=1+\frac{1}{2}+\frac{1}{3}+\dots\frac{1}{n}$$

נסתכל על שיטה להעריך את סכום הטור:

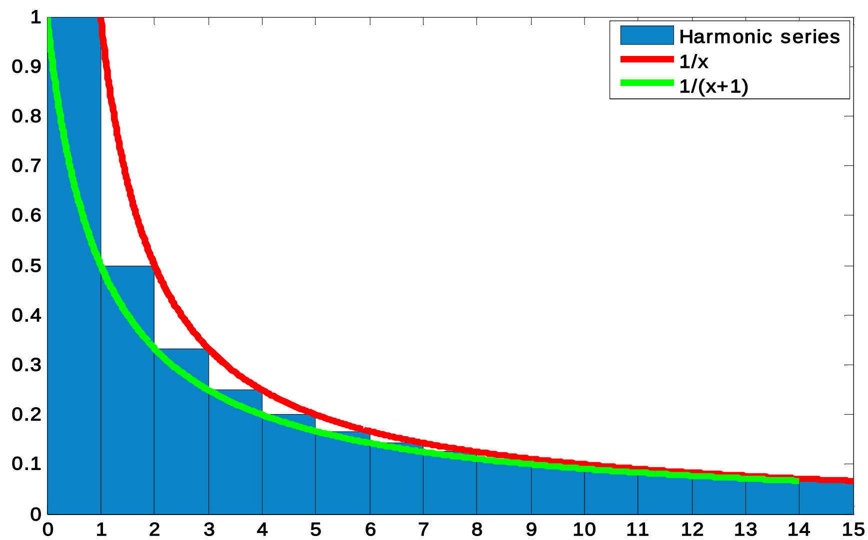
ניתן לראות את סכום הטור עד האינדקס ה-n כשטח העמודות מהעמודה הראשונה ועד העמודה ה-n.



כעת נשרטט שתי פונקציות עזר:

- $y=\frac{1}{x}$ - קו הנוגע בקצוות הימניים של העמודות (הקו האדום בציור).
- $y=\frac{1}{x+1}$ - קו הנוגע בקצוות השמאליים של העמודות (הקו הירוק בציור).

ניתן לראות המחשה לקירוב בגרף הבא:



שטח העמודות חסום בין השטחים שמתחת לגרף הפונקציות הללו:

- השטח שמתחת לקו האדום: $H_n \le 1+\int_1^n \frac{dx}{x} = 1+\ln n$
- השטח מתחת לקו הירוק: $H_n \ge H_n \ge \int_0^n \frac{dx}{x+1} = \ln(n+1)$

לסיכום, ניתן לתת חסם הדוק לסכום הטור:

$$\boxed{\ln(n+1) \le H_n \le 1+\ln n}$$

כעת ננסה להעריך את צורת הקירוב של האלגוריתם:

נסמן – ALG – פלט של האלגוריתם החמדן.

OPT – פלט של האלגוריתם האופטימאלי.

<u>משפט:</u>

$$l = \max\{|S| \in C\}$$

$$\cos t(ALG) \le H_l \bullet \cos t(OPT)$$

כאשר כזכור - $H_l \le \ln l + 1$

דוגמה (להמחשת הפרמטרים השונים) :

$$C = \{\{1,4,8\},\{3,5,7,11\},\{6,9\},\{10\}\}$$

$$Z = \{1,3,4,5,6,7,8,9,10,11\}$$

$$m = 4 \quad (Number\ of\ groups)$$

$$n = 10 \quad (Number\ \ of\ different\ elements)$$

$$l = 4 \quad (Size\ of\ the\ \mathrm{largest}\ group)$$

הוכחת המשפט (Chvatal 79'):

נשתמש בשיטת "חיוב" מיוחדת. כל לקיחת קבוצה – כמוה כתשלום על כל אלמנט בה.

הגדרה: לכל אלמנט x נגדיר $paid(x)$ לפי הזמן והקבצה שבה הוא כוסה ע"י האלגוריתם החמדן. אם כוסה ע"י קבוצה S בזמן שקבוצת האלמנטים הלא מכוסים היא X, התשלום מוגדר באופן הבא:

$$paid(x) = \frac{\mathrm{cost(S)}}{|S \cap X|}$$

נשתמש באבחנה הבאה: $\mathrm{cost(ALG)}=\sum_{x \in Z} paid(x)$ - כלומר שמבצעים פיצול של החיוב הכולל לפי אלמנטים.

למת המפתח: לכל קבוצה $S \in C$ מתקיים : $\sum_{x \in S} paid(x) \le H_{|S|} \bullet \cos t(S)$

דוגמה לקיום האבחנה הלמה:

$C = \left\{ \overset{10}{\{1,2,3,4\}}, \overset{10}{\{3.5.6\}} \right\}$ בעלת שתי תתי קבוצות בשווי 10.

האלגוריתם החמדן יבחר בתת הקבוצה הראשונה (משקל סגולי זול יותר לכל אלמנט בתת הקבוצה) ונקבל:

$paid(1) = paid(2) = paid(3) = paid(4) = \frac{10}{4} = 2.5$.

כאשר נוסיף את הקבוצה השניה נקבל: $paid(5) = paid(6) = \frac{10}{2} = 5$ - משום שיש רק שני איברים חדשים בתת הקבוצה הזו.

קיום האבחנה: $\text{cost(ALG)}=\sum_{x \in Z} paid(x) = 4 \bullet 2.5 + 2 \bullet 5 = 20$.

קיום למת המפתח: $\sum_{x \in S_1} paid(x) = 10 \le H_{|S|=4} \bullet \cos t(S_1) = (1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}) \bullet 10 = 20.8$

הוכחת למת המפתח:

נסדר את איברי S על פי הסדר בו האלגוריתם החמדן מוסיף אותם - $S = \{x_1, ..., x_d\}$.

נסתכל על האיבר הראשון x_1, לגביו מתקיים $paid(x_1) \le \frac{\text{cost(S)}}{d}$ (שיוויון עבור המקרה בו לקחנו את S). באותו אופן מתקיים עבור האיבר השני $paid(x_2) \le \frac{\cos t(S)}{d-1}$ ועבור האיבר הכללי: $paid(x_i) \le \frac{\cos t(S)}{d-i+1}$

נסכם את מחירי כל האלמנטים ונקבל את אי-השיוויון הבא:

$$\sum_{i=1}^{d} paid(x_i) \le \cos t(S) \bullet \sum_{i=1}^{d} \frac{1}{d-i+1} = \cos t(S) \bullet H_d$$

שמוכיח את למת המפתח.

הוכחת המשפט (תוך שימוש בלמה ובאבחנה):

$$\text{cost}(OPT) = \sum_{S \in OPT} \text{cost}(S) \underset{Lemma}{\ge} \sum_{S \in OPT} \frac{1}{H_{|S|}} \sum_{x \in S} paid(x)$$

$$\underset{H_{|S|} \le H_l \ for each S}{\ge} \frac{1}{H_l} \sum_{x \in Z} paid(x) \underset{observation}{=} \frac{1}{H_l} \text{cost}(ALG)$$

$$\Rightarrow \text{cost}(ALG) \le H_l \text{cost}(OPT)$$

וכך הוכחנו שהאלגוריתם החמדן נותן לכל הפחות קירוב לוגריתמי לOPT.

כעת נעבור לבחון בעיה אחרת:

<u>בעיית כיסוי צמתים (Vertex Coverage):</u>

<u>קלט:</u> גרף לא מכוון $G=(V,E)$.

<u>פלט:</u> כיסוי $V'\in V$ של כל הקשתות.

<u>המטרה:</u> למזער את גודל הכיסוי.

ראינו בשיעור הקודם אלגוריתם חמדן שנותן קירוב לוגריתמי לבעיית Set Cover. כעת ננתח דוגמה בה האלגוריתם החמדן לפתרון Set Cover מיושם לפתרון VC ונכשל:

נבנה גרף דו-צדדי:

בצד A – n צמתים.

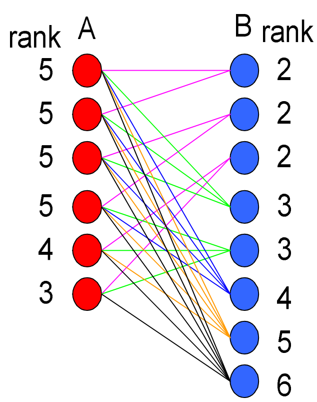
בצד B – נבנה בהדרגה:

- $\left\lfloor \frac{n}{2} \right\rfloor$ צמתים , כל אחד מחובר ל-2 צמתים ב-A.
- $\left\lfloor \frac{n}{3} \right\rfloor$ צמתים , כל אחד מחובר ל-3 צמתים ב-A.
- $\left\lfloor \frac{n}{4} \right\rfloor$ צמתים , כל אחד מחובר ל-4 צמתים ב-A.
- $\left\lfloor \frac{n}{4} \right\rfloor$ צמתים , כל אחד מחובר ל-4 צמתים ב-A.

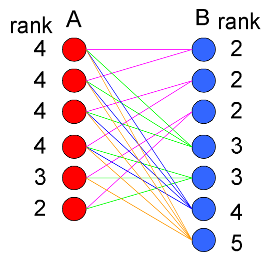
וכך הלאה עד שנגיע לצומת אחת ב-B שתחובר לכל צמתי A.

עבור $|A|=6$ יתקבל הגרף הבא:

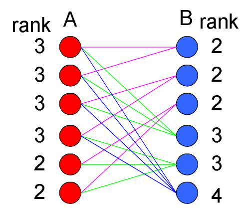
$$\boxed{\begin{aligned}|B|&=\sum_{i=2}^{n}\left\lfloor \frac{n}{i}\right\rfloor \ge \sum_{i=2}^{n}\left\lfloor \frac{n}{i}-1\right\rfloor=\sum_{i=2}^{n}\frac{n}{i}-(n-1)=\\&=n(H_n-1)-n+1=\theta(n\ln n)\end{aligned}}$$



האלגוריתם החמדן יבחר בצומת מדרגה n (6), וימחק את כל הקשתות מאותו צומת:



לאחר מכן יבחר בצומת מדרגה n-1 ושוב ימחק אותו ואת הקשתות ממנו:



ובאותה הדרך, ייבחרו כל צמתי B ראשונים, ונקבל כיסוי בגודל $\theta(n\ln n)$, כאשר OPT היה בוחר בדיוק את A בגודל n. רואים כי במקרה הגרוע מתקיים קירוב לוגריתמי ע"י האלגוריתם החמדן.

אם כך האלגוריתם החמדן לפתרון Set Cover לא מספיק טוב לפתרון בעיית VC(ראינו בשיעור הקודם כי קיים אלגוריתם חמדן לפתרון VC שנותן קירוב 2).

שאלה : איך פותרים VC עם מחיר (המטרה כעת היא מזעור מחיר הכיסוי) ?

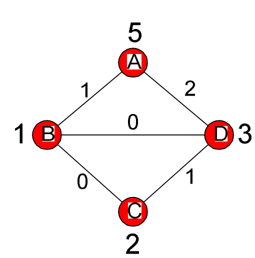
ישנו אלגוריתם הפותר עדיין את הבעייה אך שומר על קירוב 2.

נגדיר תחילה את המושג "תשלום הוגן":

לכל קשת $e \in E$ נצמיד תשלום $p(e) \geq 0$.

$p(e)$ נקרא תשלום הוגן אם לכל $v \in V$ מתקיים : $\sum_{e:(v,u)} p(e) \leq \cos t(v)$, כאשר במקרה של שיוויון נאמר שהצומת רווי.

נתבונן בדוגמה הבאה (כל התשלומים הוגנים):

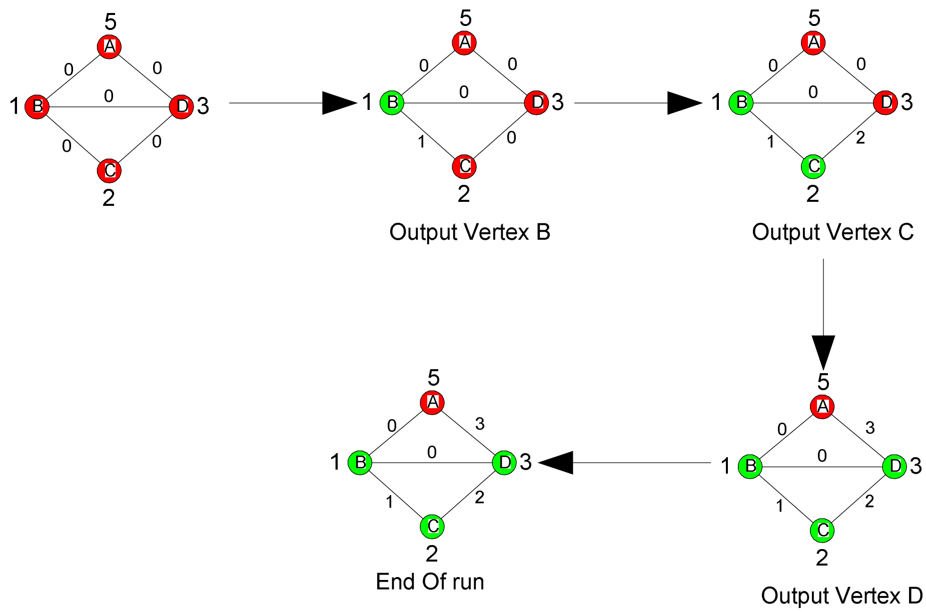


נשים לב כי בדוגמה זו הצמתים B ו D הם רווים.

תיאור האלגוריתם:

$$
\begin{array}{l}
1.\, p(e) \longleftarrow 0,\ for\ each \quad e \in E \\
2.\, find\,(e \in with \quad both \quad sides \quad unsaturated) \\
3. increase \quad p(e) \quad until \quad at \quad least \quad one \quad side \quad is \quad saturated \\
4. add \quad saturated \quad nodes \quad to \quad cover \\
5.\mathrm{goto}\, 2
\end{array}
$$

דוגמה לריצת האלגוריתם על הגרף מהדוגמה הקודמת:



משפט: האלגוריתם מבטיח מקדם קירוב 2.

הוכחה:

טענה 1: לכל כיסוי צמתים C ולכל תשלום הוגן p מתקיים: $\sum_{v\in C}\cos t(v) \ge \sum_{e\in E} p(e)$.

הוכחה:

מהגינות התשלום נובע - $\sum_{v\in C}\cos t(v) \ge \sum_{v\in C}\sum_{e:(v,u)} p(e)$

מכיוון ש-C כיסוי נובע - $\sum_{v\in C}\cos t(v) \ge \sum_{v\in C}\sum_{e:(v,u)} p(e) \ge \sum_{e\in C} p(e)$ - כי C נוגע בכל הקשתות.

טענה 2: עבור התשלום הסופי p מתקיים : $\cos t(ALG) \le 2\sum_{e\in E} p(e)$.

הוכחה (אינטואיטיבית) : נחשוב שהקשתות "משלמות" על הצמתים, אזי כל קשת משלמת לכל היותר פעמיים – פעם לכל קצה.

הוכחת המשפט:

התשלום p הסופי הוא הוגן לפי האלגוריתם, לכן $\cos t(ALG) \underset{claim2}{\le} 2\sum_{e\in E} p(e) \le 2\sum_{v\in OPT} \cos t(v)$

מכאן נובע ישירות כי: $\cos t(ALG)\le 2\cos t(OPT)$.

ניתן לראות באלגוריתם מאין בעיית מילוי מים, כאשר הקשתות "מתמלאות" לכל היותר עד לאיזון רמת המים בין הצמתים שלהם.

<u>בעיית k – מרכזים (K-Center):</u>

<u>קלט:</u> גרף ממושקל $G=(V,E,W)$ ומספר k

<u>פלט:</u> תת קבוצה C של k צמתים.

<u>**המטרה:**</u> למזער את $\max_{v\in V}\{dist(C,v)\}$.

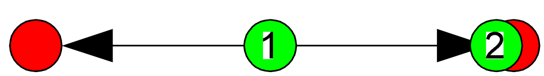
הגדרה: לקבוצה C של צמתים ($C\le V$) ולצומת $v\in V$ מוגדר מרחק מהצומת למרכז השירות הקרוב:

$dist(C,v)=\min_{u\in C}(dist(u,v))$.

אינטואיציה ראשונה לפתרון הבעיה – אלגוריתם חמדן. בכל פעם נעבור על הגרף ונבחר צומת שימזער את המרחק המקסימאלי.

<u>אלגוריתם חמדן נכשל כאן ובגדול!!</u>

נסתכל על דוגמה בה ניתן למקם מרכזים לא רק בצמתים אלא בכל מקום שנבחר:



עבור k=2 ימקם החמדן מרכז ראשון בדיוק באמצע, ואת השני ליד אחד הצדדים. כך נקבל מרחק גדול מהצד בו לא ימוקם מרכז.

נבחן אלגוריתם לקירוב הפתרון:

$$
\begin{aligned}
&C \xleftarrow{any} \{v_1\}\\
&for\ i=2\ \ to\ \ k\ \ do\\
&\qquad x_i \longleftarrow \arg\max_{v\in V}\{dist(C,v)\}\\
&\qquad C \longleftarrow C\bigcup\{x_i\}\\
&end\\
&output\ \ C
\end{aligned}
$$

משפט: האלגוריתם הוא 2 קירוב.

הוכחה:
נסמן $r_i = \max(dist(v,\{x_1,...x_i\}))$. בסיום הריצה ישנם $\{x_1,...x_k\}$ מרכזים והם מכסים את כל האיברים ב"כדורים" ברדיוס r_k.

נסתכל על פתרון OPT עם רדיוס כיסוי r^*. נביט על הכדורים ברדיוס r^* סביב המרכזים של OPT.
נבחין בין 2 מקרים:

1. בכל כדור כזה יש צרכז של ALG – אם כך מתקיים אי שיוויון משלוש:
$dist(v,ALG) \le dist(v,OPT_i) + dist(OPT_i, ALG_i) \le 2r^* \;<\; dist(v,ALG) \le 2r^*$

2. יש כדור של OPT ללא מרכז של ALG. מכאן שעל פי עקרון שובח היונים ישנו כדור של OPT המכיל שני מרכזים של ALG. ALG בוחר בכל פעם את הצומת הרחוק ביותר ולכן המרחק בין שני מרכזים הוא לכל היותר $2r^*$ (על קוטר הכדור במרכזו מרכז של OPT). בנוסף, המרחק בין כל שתי נקודות של ALG הוא לפחות r_k (רדיוס הכיסוי של ALG). מכאן מקבלים $r_k \le 2r^*$.