14: Approximating min VC

נזכיר: בגרף G, קבוצה $C \subseteq V(G)$ תיקרא **כיסוי בקודקודים** של G אם לכל צלע $uv \in E(G)$, מתקיים $u \in C$ או $v \in C$.

אלגוריתם אקראי

1. נאתחל $C = \emptyset$.
2. כל עוד $E \neq \emptyset$, נבצע:
   a. נבחר $e = uv \in E$ שרירותית.
   b. $C \coloneqq C + v$
   c. $E \coloneqq E \setminus \{e \in E : v \in e\}$. נוריד מ-$E$ את כל הצלעות שמחוברות ל-v.
3. נחזיר את C.

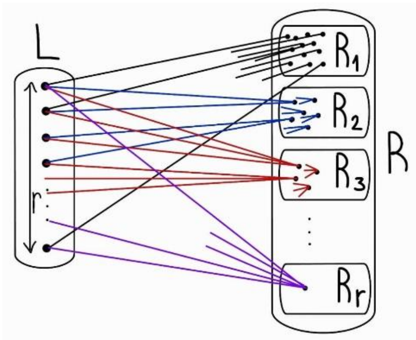
נשקול את הביצועים של האלגוריתם על הגרף הדו"צ הבא:

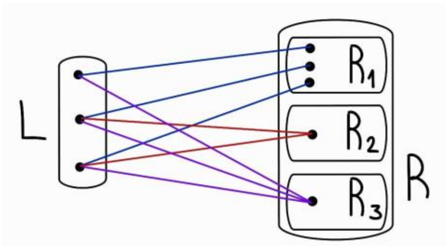
$$G = (L \cup R, E), \qquad R \coloneqq R_1 \cup R_2 \cup \ldots \cup R_r, \qquad r = |L|, \qquad \forall i : |R_i| = \lfloor r/i \rfloor$$

כלומר: יש r קודקודים ב-L, ו-R מחולקת ל-r קבוצות. הגודל של כל קבוצה i הוא r/i (מעוגל כלפי מטה).

בנוסף, בתוך כל R_i, אין שני קודקודים עם שכן משותף. ולכל קודקוד ב-R_i, יש i שכנים ב-L.

המבנה הכללי: ודוגמה עבור $r = 3$: $|R_1| = 3/1 = 3,\ |R_2| = \lfloor 3/2 \rfloor = 1,\ |R_1| = \lfloor 3/3 \rfloor = 1$





כמה צלעות יהיו בגרף? נבדוק כמה קודקודים יהיו ב-R. לפחות:

$$\sum_{i=1}^{r} \lfloor r/i \rfloor = \Theta\left(r \cdot \sum_{i=1}^{r} 1/i\right) = \Theta(r \cdot \log r)$$

(חסם על הטור ההרמוני). מספר הקודקודים ב-R הוא חסם תחתון למספר הצלעות שיש בגרף. אז מספר הצלעות הוא ב- $\Omega(r \cdot \log r)$.

בגרף שלנו, L מהווה פתרון VC אופטימלי יחיד (אין עוד פתרון מינימום אחר).

אבל האלגוריתם יכול להחזיר את R (אם בכל פעם שהוא בוחר צלע שרירותית, הוא בוחר קודקוד מ-R). ואז, יחס הקירוב הוא:

$$\frac{|R|}{|L|} = \frac{\Omega(r \cdot \log r)}{r} = \Omega(\log r)$$

תלוי בגודל הגרף. זה יחס קירוב רע מאוד (כי r הוא פונקצייה של n).

אלגוריתם חמדן

האלגוריתם כמעט זהה לקודם, אבל במקום לבחור שרירותית, נבחר בצורה חמדנית:

1. נאתחל $C = \emptyset$.
2. כל עוד $E \neq \emptyset$, נבצע:
   a. נבחר $v \in V$ בעל דרגה מקסימום (אם יש יותר מאחד, נבחר ביניהם שרירותית).
   b. $C \coloneqq C + v$
   c. $E \coloneqq E \setminus \{e \in E : v \in e\}$. נוריד מ-$E$ את כל הצלעות שמחוברות ל-v.
3. נחזיר את C.

כל פעם נבחר קודקוד שיש לו כמה שיותר צלעות, אז ברעיון נכסה את כל הצלעות עם פחות קודקודים.

אבל עדיין, בגרף שהראינו, אפשר לגרום לאלגוריתם להחזיר את R. כי בבחירה הראשונה נוכל לבחור את הקודקוד ב-R_r,

ואז זה מוריד צלע מכל אחד מהקודקודים של L. ובכל פעם שבחרנו קודקוד מ-R_i כלשהו, לא הורדנו צלעות מאף R_j אחר.

אז בכל פעם נוכל לבחור קודקוד מ-R.

אלגוריתם 2-מקרב, מבוסס על שידוך מקסימום

ניזכר ששידוך מקסימום היא בעיה ב-P. וגם, שגודל הכיסוי המינימום $\tau(G)$ הוא לפחות גודל השידוך המקסימום $\nu(G)$.

אז נוכל למצוא שידוך מקסימום בגרף, ולקחת את הקודקודים של הצלעות של השידוך.

כי זה בוודאות מכסה את כל הצלעות (אם יש צלע שלא מכוסה, אפשר להוסיף אותה לשידוך).

מספר הצלעות הוא לכל היותר כגודל הכיסוי ($\nu(G) \leq \tau(G)$), ולקחנו 2 קודקודים לכל צלע.

כלומר לקחנו לכל היותר $2 \cdot \tau(G)$ קודקודים. פי 2 מגודל הכיסוי המינימום, כלומר 2-קירוב.

הבעיה באלגוריתם הזה, היא שהוא דורש את אלגוריתם *Edmonds* (מציאת שידוך מקסימום בגרף כללי), שהוא מסובך ויקר בזמן ריצה. (פולינומי, אבל עדיין לא משהו).

Savage's Algorithm **– אלגוריתם 2-מקרב בזמן לינארי**

<u>קלט</u>: גרף G קשיר. (אם הוא לא קשיר, נפעיל את האלגוריתם על כל רכיב קשירות בנפרד).

1. יהי T עץ DFS של G.
2. נחזיר את הקודקודים הפנימיים של T (הכל חוץ מהעלים).

באופן טריוויאלי, כל הצלעות מכוסות. כי אין צלע בין שני עלים: כשביקרנו בעלה הראשון מביניהם, אם הייתה צלע היינו עוברים בה (ואז זה לא היה עלה).

כדי להוכיח שהוא גם 2-מקרב, מספיק להוכיח:

למה: יהי T עץ DFS של גרף קשיר, ונסמן $I(T)$ את הקודקודים הפנימיים. אזי, T מכיל שידוך בגודל לפחות $|I(T)|/2$.

אם הלמה נכונה, אז נקבל ש $\tau(G) \geq |I(T)|/2$, כי הכיסוי צריך לקחת לפחות קודקוד אחד מכל צלע בשידוך.

האלגוריתם מחזיר קבוצה בגודל $|I(T)|$, אז הוא 2-מקרב.

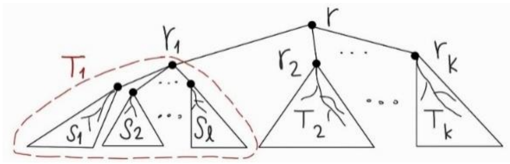
נוכיח את הלמה: באינדוקציה על כמות הצלעות. (נתעלם מעצים שהם רק קודקוד יחיד, כי אין צלעות).

<u>בסיס</u>: עבור $e(T) = 1$, נקבל: $|I(T)| = 0$, ואכן T מכיל שידוך בגודל $0 \leq 1$.

<u>צעד</u>: נניח שהטענה מתקיימת עבור כל תת-עץ ממש של T, ויהי T עץ DFS של גרף קשיר.

נסמן r את שורש העץ. ונסמן $T_1, T_2, \dots, T_k$ את תתי-העצים המושרשים בילדים של r. r_i הוא השורש של T_i.

נסמן $S_1, S_2, \dots, S_\ell$ את תתי-העצים של T_1 (מושרשים ב- r_1):



נחשב את גודל השידוך שיש ב-T:

בכל S_i, יש שידוך בגודל לפחות $|I(S_i)|/2$ (לפי הנ"א). כלומר ב- T_1 יש שידוך בגודל לפחות:

$$\frac{1}{2} \cdot \sum_{i=1}^{\ell} |I(S_i)|$$

בכל $T_2, T_3, \dots, T_k$ יש שידוך בגודל לפחות $|I(T_i)|/2$ (לפי הנ"א). כלומר ב- $T_2 \cup T_3 \cup \dots \cup T_k$ יש שידוך בגודל לפחות:

$$\frac{1}{2} \cdot \sum_{i=2}^{k} |I(T_i)|$$

ואנחנו יודעים שלא השתמשנו ב r או r_1. אז נוסיף את rr_1 לשידוך. בסה"כ קיבלנו שיש שידוך ב-T בגודל לפחות:

$$\geq 1+\frac{1}{2}\cdot\sum_{i=1}^{\ell}|I(S_i)|+\frac{1}{2}\cdot\sum_{i=2}^{k}|I(T_i)|=1+\frac{\sum_{i=1}^{\ell}|I(S_i)|+\sum_{i=2}^{k}|I(T_i)|}{2}$$

ונשים לב ש: $\sum_{i=1}^{\ell}|I(S_i)|+\sum_{i=2}^{k}|I(T_i)|$ זה מספר הקודקודים הפנימיים בעץ, בלי r, r_1. אז:

$$\sum_{i=1}^{\ell}|I(S_i)|+\sum_{i=2}^{k}|I(T_i)|=|I(T)|-2$$

נציב את השוויון הזה בחישוב, ונקבל שב-T יש שידוך בגודל לפחות:

$$\geq 1+\frac{|I(T)|-2}{2}=\frac{2+|I(T)|-2}{2}=\frac{|I(T)|}{2}$$

כנדרש.

Minimum Weight Vertex Cover

בהינתן גרף G, **ופונקציית משקל** $w: V(G)\to\mathbb{R}^+$, על הקודקודים של G. נרצה למצוא כיסוי $C\subseteq V(G)$ שממזער את $\sum_{v\in C}w(v)$.

אף אחד מהאלגוריתמים שראינו לא עובד עם משקלים. באופן כללי, כי משקלים לא קשורים למבנה הגרף. נצטרך להיעזר בתכנון לינארי.

נתחיל ב- *Integer Programming*. נקודד את *min-w-VC* לבעיית *IP*, שהיא *NPC*:

יהיה לנו וקטור x שמייצג איזה קודקודים לוקחים לכיסוי. המיקום ה-v ב-x יהיה 1 אם $v\in C$, ו-0 אחרת. ואנחנו נרצה למזער את:

$$\sum_{v\in V(G)}w(v)\cdot x_v$$

תחת האילוצים: $x_u+x_v\geq 1$ לכל $uv\in E(G)$, ו- $x_v\in\{0,1\}$ לכל v.

נעבור ל-*relaxation* של הבעיה ל-*LP*. נקודד את *min-w-VC* לבעיית *LP*, שהיא *P*:

הכל אותו דבר, אבל האילוץ של x הופך להיות $x_v\geq 0$.

יכולנו לרשום $0\leq x_v\leq 1$, אבל זה אותו דבר.
כי אנחנו נרצה למזער את הסכום שתלוי ב-x, וברגע שיש יותר מ-1 הקודקוד כבר מכוסה (אז היה אפשר להגיע לאותו כיסוי עם 1).

האלגוריתמים *Simplex*, *Ellipsoid* יכולים למצוא פתרון אופטימלי בזמן פולינומי.

אבל איך יודעים שהפתרון שנמצא ב-*LP* פותר את הבעיה שתיארנו ב-*IP*?

נציע אלגוריתם שמבוסס על *LP*, למציאת כיסוי בקודקודים במשקל מינימום:

1. יהי x^* הפיתרון האופטימלי לבעיית ה-*LP*.
2. נחזיר: $C\coloneqq\{v\in V(G) : x_v^*\geq 0.5\}$.

טענה: C היא אכן כיסוי בקודקודים.

הוכחה: יהי $uv\in E(G)$. אזי:

$$x_v^*+x_u^*\geq 1 \implies x_v^*\geq 0.5 \;\; or \;\; x_u^*\geq 0.5$$

כלומר, בפיתרון של ה-*LP* דרשנו $x_v^*+x_u^*\geq 1$ לכל uv. ואם התנאי הזה מתקיים, אז אחד הקודקודים קיבל לפחות חצי.
כלומר נבחר לכיסוי לפחות קודקוד אחד מכל צלע, כנדרש.

עכשיו שהראנו שהפתרון הוא אכן כיסוי בקודקודים, נראה עד כמה הוא מתרחק מהמשקל המינימום (האופטימלי).

טענה: האלגוריתם הוא 2-מקרב.

הוכחה: נסמן *OPT* את המשקל האופטימלי של הפתרון ל-*IP*, ו-*OPTf* את המשקל האופטימלי של הפתרון ל-*LP*.

האילוץ $0\leq x_v\leq 1$ גורר ש: $OPTf\leq OPT$. כי אותם קודקודים ייבחרו, ובגרסת ה-*LP* יתקיים $0\leq x_v\leq 1$ כמו שאמרנו.

לכל קודקוד שנמצא ב-C, היה משקל לפחות חצי ב- x^*. כלומר מתקיים:

$$\sum_{v\in C}\frac{w(v)}{2}\le\sum_{v\in v(G)}w(v)\cdot x_v^*$$

כלומר:

$$\sum_{v\in C}w(v)\le 2\cdot\underbrace{\sum_{v\in v(G)}w(v)\cdot x_v^*}_{OPTf}=2\cdot OPTf\le 2\cdot OPT$$

כנדרש.

אלגוריתם מקרי לכיסוי קודקודים מינימום (בלי משקלים)

קלט: גרף $G\coloneqq(V,E)$

1. נאתחל $C=\emptyset$.
2. נסדר את E בסדר שרירותי.
3. כל עוד $E\ne\emptyset$, נבצע:
   a. נבחר את ה- $e=uv\in E$ הבא, לפי הסדר שקבענו.
   b. נבחר $x\in\{u,v\}$ באופן מקרי ואחיד.
   c. $C\coloneqq C+x$
   d. $E\coloneqq E\setminus\{e\in E:x\in e\}$.
4. נחזיר את C.

האלגוריתם תמיד יחזיר VC תקין, כי מורידים צלע רק אם בחרנו קודקוד שנוגע בה.

טענה: התוחלת של יחס הקירוב היא 2. (כלומר נצפה שהאלגוריתם יהיה 2-מקרב, שהאלגוריתם מחזיר כיסוי בגודל לכל היותר פעמיים הגודל האופטימלי).

הוכחה: נסמן C^* כיסוי אופטימלי כלשהו. נסמן C את הכיסוי שקיבלנו מהאלגוריתם. אז $c\coloneqq|C|$ הוא משתנה מקרי. נוכיח ש: $\mathbb{E}[c]\le 2|C^*|$.

האלגוריתם מטיל סה"כ c מטבעות הוגנים. כל הטלה תקבע האם הקודקוד ייכנס ל-C.

הטלת מטבע תיקרא **טובה** אם הקודקוד שהכנסנו נמצא גם ב- C^*.

ניזכר שלכל צלע uv, לפחות אחד מתוך הקודקודים נמצא ב- C^*. כלומר לכל זריקה יש הסתברות לפחות חצי להיות טובה.

מספר ההטלות הטובות חסום ב- $|C^*|$. (כי לא נוכל לבחור יותר מ- $|C^*|$ קודקודים ששייכים ל- C^*. אחרי $|C^*|$ הטלות טובות, נקבל $C^*\subseteq C$).

נסמן X את מספר ההטלות שנדרשות כדי לקבל $|C^*|$ הטלות טובות. (זה גם מ"מ).

אזי, $c\le X$. כי ברגע ש- $c=X$, זה אומר שהטלנו X פעמים, אז קיבלנו $|C^*|$ הטלות טובות, יתקיים $C^*\subseteq C$ ונסיים.

אז גם $\mathbb{E}[c]\le\mathbb{E}[X]$. אם נראה ש- $\mathbb{E}[X]\le 2|C^*|$, זה יוכיח את הטענה.

נסמן X_1 את מספר ההטלות לפני ההטלה הטובה הראשונה. ונסמן X_2 את מספר ההטלות מההטלה הטובה הראשונה, עד ההטלה הטובה השנייה.

באופן כללי: X_i זה מספר ההטלות בין ההטלה הטובה ה- $i-1$ ועד ההטלה הטובה ה- i.

אז מתקיים: $X=X_1+X_2+\cdots+X_{|C^*|}$, כי אחרי $|C^*|$ הטלות נקבל $C^*\subseteq C$.

מלינאריות התוחלת, מתקיים:

$$\mathbb{E}[X]=\sum_{i=1}^{|C^*|}\mathbb{E}[X_i]$$

בנוסף, לכל i: $X_i\sim Geo(0.5)$. אז $\mathbb{E}[X_i]=1/0.5=2$. (תוחלת של $Geo(p)$ זה $1/p$).

כלומר, $\mathbb{E}[X]=2\cdot|C^*|$, כנדרש.