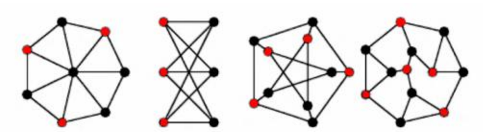
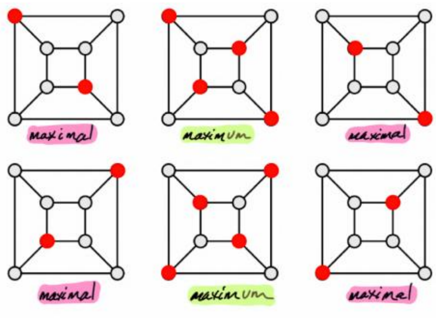
Independent Sets in Graphs

יהי גרף G. קבוצת קודקודים $I \subseteq V(G)$ תיקרא בלתי תלויה אם היא לא פורשת אף צלע (אין אף צלע בין כל שני קודקודים בקבוצה).



נסמן $\alpha(G)$ את גודל הקבוצה הבת"ל הגדולה ביותר (מקסימום).

מקסימום זאת הקבוצה הכי גדולה שאפשר למצוא. **מקסימלי** זו קבוצה שלא ניתן להגדיל. כל מקסימום הוא מקסימלי, ולא בהכרח הפוך.



הבעיה של מציאת קבוצה מקסימלית היא ב-P. נפעיל אלגוריתם חמדן: נבחר קודקוד ונוריד את השכנים שלו, ככה לכל הקודקודים.

הבעיה של מציאת קבוצת מקסימום היא NPC. נגדיר את השפה:

$$IS := \{(G,k) : \alpha(G) \geq k\}$$

נוכיח שהיא NPC. **תחילה נוכיח שהיא ב-NP.** נגדיר אלגוריתם אימות:

קלט: גרף G, קבוצת קודקודים $V' \subseteq V(G)$. לכל $u,v \in V'$, אם קיימת צלע uv בגרף, נחזיר 0. אם בדקנו את כל הזוגות ואין צלע, נחזיר 1.

האלגוריתם מקיים את התנאים (באופן טריוויאלי):

1. האלגוריתם מחזיר 1 אם ורק אם $(G,|V'|) \in IS$.
2. הגודל של V' פולינומי ב- $|V|$.
3. האלגוריתם רץ בזמן $O(|G| + |V'|^2)$.

נוכיח שהשפה NPH: נראה ש $IS \leq_p$ 3-CNF-SAT. כלומר נראה רדוקציה - פונקציה f כך שבהינתן נוסחה φ שהיא $3\text{-}CNF$, יתקיים:

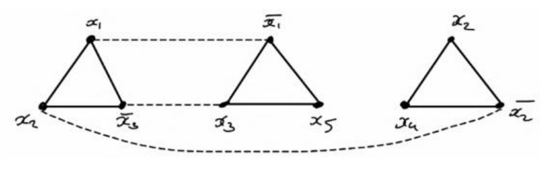
1. $f(\varphi)$ הוא זוג סדור (G_φ, k).
2. $\alpha(G_\varphi) \geq k$ **אם ורק אם** φ ספיקה.
3. ניתן לחשב את $f(\varphi)$ בזמן פולינומי.

נגדיר את f: בהינתן φ נוסחת $3\text{-}CNF$ עם m פסוקיות, נגדיר גרף G_φ כך:

לכל פסוקית $C := \ell_1 \vee \ell_2 \vee \ell_3$ ב-φ נגדיר משולש עם קודקודים שהם הליטרלים.

בנוסף, קודקודים שנמצאים על משולשים שונים יתחברו בצלע אם הליטרלים שלהם משלימים. לדוגמה:

עבור $\varphi = (x_1 \vee x_2 \vee \overline{x_3}) \wedge (\overline{x_1} \vee x_5 \vee x_3) \wedge (x_4 \vee \overline{x_2} \vee x_2)$, נקבל:



הפונקציה f מחזירה את (G_φ, m). נוכיח שהיא עומדת בתנאים:

1. מתקיים לפי הבנייה.
3. מספר המשולשים בתחילת הבנייה שווה למספר הפסוקיות ב-φ. מספר הצלעות גם פולינומי בגודל של φ, כי הוא חסום ב- $(3m)^2$.

נוכיח את 2:

כיוון ראשון: נניח ש φ ספיקה, ותהי a השמה מספקת כלשהי עבור φ. כלומר a מספקת לפחות ליטרל אחד בכל פסוקית.

נבחר ליטרל מסופק אחד (שרירותית) מכל פסוקית. כל ליטרל כזה מתאים לקודקוד בגרף.

קבוצת הקודקודים הזו היא קבוצה בת"ל בגודל m בגרף G_φ. למה?

בחרנו ליטרל אחד מכל פסוקית, אז זה נותן לנו בדיוק m קודקודים. נב"ש שהקבוצה תלויה, כלומר שקיימים שני קודקודים שיש ביניהם צלע.
לפי הגדרת הבנייה, זה קורה אמ"מ הם משלימים. אבל בחרנו ליטרלים מסופקים, אז לא יכול להיות שבחרנו את שניהם. מש"ל.

כיוון שני: נניח ש $\alpha(G_\varphi) \geq m$, ותהי $I \subseteq V(G_\varphi)$ קבוצה בת"ל בגודל m ב-G_φ.
הקבוצה I חייבת להכיל בדיוק קודקוד אחד מכל משולש (כי אם יש 2 ממשולש, הקבוצה לא בת"ל. וצריך m קודקודים).
כלומר: לכל משולש ב-G_φ, יש בדיוק קודקוד אחד ששייך ל-I.
נגדיר השמה a' שמספקת את φ, בצורה הבאה:
כל ליטרל שנמצא ב-I, יסופק (יקבל ערך $TRUE$). עבור שאר המשתנים, ניתן השמה שרירותית, שאינה סותרת.
אז לכל פסוקית יש לפחות ליטרל אחד שקיבל $TRUE$. ואין סתירות, כי אין ליטרלים ב-I שמופיעים גם המקורי וגם המשלים ב-φ.
(כי אחרת הייתה ביניהם צלע, וזו סתירה לכך ש-I בת"ל).
בסה"כ, הוכחנו $IS \in NP \cap NPH = NPC$.

Graph Coloring

k-coloring (k-צביעה): עבור גרף G, פונקציה $f: V(G) \to [k]$ תיקרא k-צביעה של G אם לכל $(u,v) \in E(G)$, מתקיים $f(u) \neq f(v)$.
Chromatic Number (מספר כרומטי): נסמן ב- $\chi(G)$ את ה-k הכי קטן שעבורו קיימת k-צביעה של G.
אבחנה: עבור k-צביעה כלשהי, כל "מחלקת צבע" (קבוצה של קודקודים שכולם באותו הצבע, $\{v \in V(G): f(v) = i\}$) היא קבוצה בת"ל.
(לפי הגדרת k-צביעה). לכן, k-צביעה של גרף שקולה לחלוקת הקודקודים ל-k קבוצות בת"ל.
גרף G הוא 2-צביע אמ"מ הוא דו-צדדי (*Bipartite*). שאפשר לחלק את הקודקודים שלו ל-2 קבוצות שאין ביניהן צלע.
נגדיר את השפות:

$$2COL \coloneqq \{G: \chi(G) \leq 2\}, \qquad 3COL \coloneqq \{G: \chi(G) \leq 3\}$$

באופן טריוויאלי, $2COL$ היא ב-P כי זה שקול לבדיקה האם גרף הוא דו צדדי (בדיקה האם יש בגרף מעגלים אי זוגיים).
עוד אבחנה: נסמן $n = V(G)$, אזי מתקיים:

$$\alpha(G) \geq n/\chi(G)$$

גודל הקבוצה הבת"ל הכי גדולה היא לפחות מספר הקודקודים חלקי המספר הכרומטי. אם הגרף הוא 3-צביע,
אז אפשר למצוא 3 קבוצות בת"ל וזה אומר שהקבוצה הבת"ל הגדולה היא **לפחות** בגודל שליש הקודקודים.
טענה: $3COL$ היא ב-NPC. כמו עם k-CNF-SAT, המעבר מ-2 ל-3 מעביר את הבעיה מ-P ל-NPC. כדי להוכיח את זה, נציג עוד שפה:

NAE-SAT

בהינתן נוסחת k-CNF כלשהי φ, נאמר שהיא בתצורת *NAE-SAT* (או *"Not All Equal" Satisfiable*),
אם קיימת לה השמה מספקת כך שבכל פסוקית קיים לפחות ליטרל אחד לא מסופק. (מעצם ההגדרה שההשמה מספקת, יש לה לפחות גם ליטרל מסופק אחד).
נגדיר את השפה *NAE-k-CNF-SAT*, כל הנוסחאות במבנה k-CNF שהן גם *NAE-SAT*:

$$\text{NAE-k-CNF-SAT} \coloneqq \{\varphi: \varphi \text{ is an NAE-satisfiable k-CNF formula}\}$$

השפה *NAE-3CNF-SAT* היא NPC.

כעת, נוכיח ש-3-COL היא ב-NPC. קודם **נוכיח שהיא ב-*NP*** על ידי הגדרת אלגוריתם אימות:
קלט: גרף G, 3-צביעה f. לכל צלע, נבדוק אם $f(u) = f(v)$. אם זה מתקיים באחת הצלעות, נחזיר 0. אם בכל הצלעות זה לא מתקיים, נחזיר 1.
מתקיים באופן טריוויאלי: האלגוריתם מחזיר 1 אמ"מ הגרף הוא 3-צביע, הגודל של f (התחום) פולינומי בגודל של G, והאלגוריתם פולינומי ב- $|G| + |f|$.

נוכיח שהשפה ב-*NPH*: מספיק להראות ש: $\text{NAE-3-CNF-SAT} \leq_p 3COL$, ע"י רדוקציה.
נתאר פונקציה f שבהינתן נוסחה $\varphi \in \text{3-CNF}$:
1. $f(\varphi)$ הוא גרף G_φ, 2. מתקיים: $\chi(G_\varphi) \leq 3$ **אמ"מ** $\varphi \in \text{NAE-3-CNF-SAT}$, 3. ניתן לחשב את $f(\varphi)$ בזמן פולינומי.

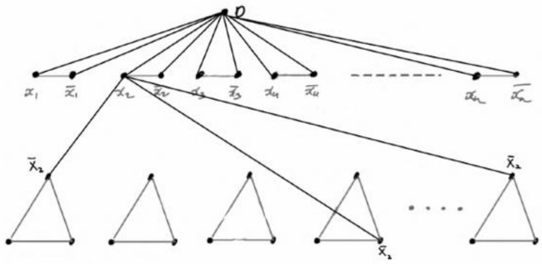
2: Classical NPC Languages

הגדרת f: בהינתן φ נוסחת *3-CNF*, נגדיר גרף G_φ כך:

לכל פסוקית $C \coloneqq \ell_1 \vee \ell_2 \vee \ell_3$ ב-φ, נגדיר משולש כמו ברדוקציה הקודמת (נקרא לזה **"הגאדג'ט של הפסוקיות"**).

בנוסף, לכל משתנה x_i ב-φ נגדיר צלע $e = (x_i, \bar{x}_i)$ (**"הגאדג'ט של המשתנים"**). ונוסיף קודקוד D (*don't care*) לגרף.

נחבר את D עם כל קודקוד בגאדג'טים של המשתנים. ונחבר כל ליטרל בגאדג'טים של המשתנים עם המשלים שלו בגאדג'טים של הפסוקיות:



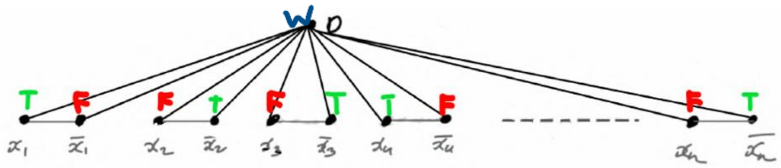
נוכיח ש-f מקיימת את התנאים:

1 טריוויאלי לפי הגדרת הבנייה, 3 מתקיים כי מספר הגאדג'טים והצלעות פולינומי במספר המשתנים ומספר הפסוקיות.

נוכיח את 2 – נראה ש $\varphi \in \text{NAE-3-CNF-SAT} \iff \chi(G) = 3$. (פורמלית היינו מוכיחים $\chi(G) \le 3$, אבל יש משולשים אז השוויון טריוויאלי).

כיוון ראשון: נניח ש φ בצורת *NAE-SAT*, ותהי a השמה מספקת *NAE* עבורה.

נגדיר 3-צביעה עבור G_φ באופן הבא: הצבעים יהיו *T, F, W*, ונצבע את הקודקודים בגאדג'טים של המשתנים בהתאם ל-a, לדוגמה:



נגדיר את הצביעה עבור הגאדג'טים של הפסוקיות (המשולשים):

עבור פסוקית $C \coloneqq \ell_1 \vee \ell_2 \vee \ell_3$, נסרוק אותה משמאל לימין ונעצור בליטרל המסופק הראשון. נצבע אותו T.

מכיוון שזו השמת *NAE*, יש ליטרל אחר שלא מסופק ב-C. נצבע אותו ב-F. ונותר עוד ליטרל אחד ב-C, נצבע אותו W.

נותר להראות שהצביעה היא 3-צביעה תקינה (שאין שני קודקודים סמוכים באותו הצבע):

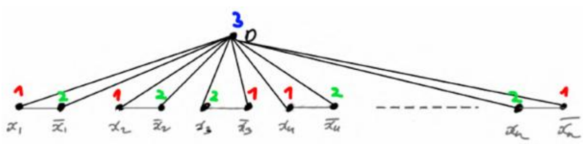
נשים לב ש-W לא בעייתי כי הוא מופיע רק פעם אחת במשולש ובקודקוד D. המשולשים לא מחוברים ישירות ל-D ואין צלעות בין המשולשים.

כמו כן, אין שני ליטרלים באותה השמה באותו משולש או בגאדג'ט של המשתנים.

אז הבעיה היחידה שיכולה להיות היא האיור הימני:

אבל כל החיבורים האלה הם רק מליטרל למשלים שלו (כמו באיור השמאלי), ולפי הצביעה לא יהיה מצב כזה ששניהם באותו צבע.

כיוון שני: נתון $\chi(G_\varphi) = 3$. תהי ψ פונקציית ה-3-צביעה. נניח בה"כ שהצבעים 1,2 מופיעים בגאדג'טים של המשתנים ולכן D נצבע ב-3.



נגדיר השמה בתצורת *NAE-SAT* שמספקת את φ: היא תהיה בהתאם לצביעה כאשר 1 זה T, 2 זה F.

נראה שההשמה תקינה – כל משתנה בעל ערך יחיד: בגאדג'טים של המשתנים כל משתנה מופיע בדיוק פעמיים – ליטרל והמשלים שלו.

פעם אחת חיובי, ופעם ושלילי. אז אין סתירות בהשמה.

נראה שהיא בתצורת *NAE-SAT*: מכיוון שהצביעה היא 3-צביעה, בכל משולש יש שימוש בכל הצבעים. ובפרט צבעים 1,2 כלומר בכל פסוקית יש *F,T*.

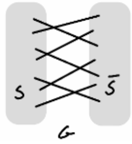
בסה"כ, הוכחנו ש-*3COL* היא *NP* וגם *NPH* ולכן *NPC*.

2: Classical NPC Languages

The Max-Cut Problem

בהינתן גרף G, קבוצת קודקודים $S \subseteq V(G)$, והקבוצה המשלימה $\bar{S} = V(G) \setminus S$, נסמן את קבוצת הצלעות החוצות מ-S ל-$\bar{S}$ כך:

$$E_G(S,\bar{S}) = \{uv \in E(G) : \ u \in S, v \in \bar{S}\}$$



החלוקה הזו נקראת **חתך** (*cut*). בנוסף, נסמן את גודל החתך: $e_G(S,\bar{S}) = |E_G(S,\bar{S})|$. גודל החתך המקסימום בגרף G:

$$\sigma(G) \coloneqq \max_{S \subseteq V(G)} e_G(S,\bar{S})$$

זוג $(S,\bar{S})$ שמקיים $e_G(S,\bar{S}) = \sigma(G)$ ייקרא **חתך מקסימום** של G.

נגדיר את השפה *MAX-CUT*, קבוצת הזוגות: $\{(G,k) : \sigma(G) \geq k\}$, גרפים שהחתך המקסימום שלהם לפחות בגודל k. טענה: היא *NPC*.

נגדיר אלגוריתם אימות כדי להראות שהיא *NP*:

קלט: גרף G, חלוקה של הקודקודים $(S,\bar{S})$, מספר טבעי k.

נגדיר $x = 0$. לכל צלע uv, אם הצלע מחברת בין הקבוצות $S,\bar{S}$, נוסיף 1 ל-x. אחרי שעברנו על הצלעות: אם $x \geq k$ נחזיר אמת. אחרת, שקר.

הנכונות של האלגוריתם טריוויאלית – זו ספירה של מספר הצלעות בחתך. הפולינומיות של הגודל של העד וזמן הריצה גם טריוויאלית.

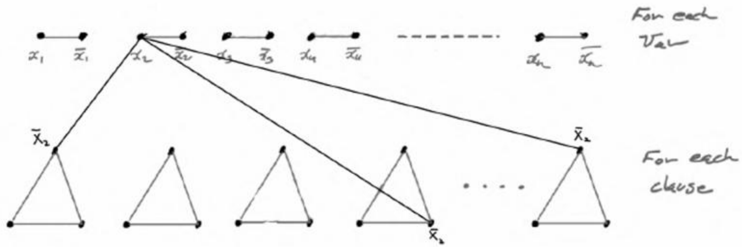
זמן הריצה הוא $O(|E(G)|)$.

כדי להראות שהשפה *NPH*, נראה ש $\text{MAX-CUT} \leq_p \text{NAE-3-CNF-SAT}$ באמצעות הפונקציה f, שצריכה לקיים:

1. $f(\varphi) = (G_\varphi, k)$,
2. $\sigma(G_\varphi) \geq k \Leftrightarrow \varphi \in \text{NAE-3-CNF-SAT}$,
3. אפשר לחשב את $f(\varphi)$ בזמן פולינומי

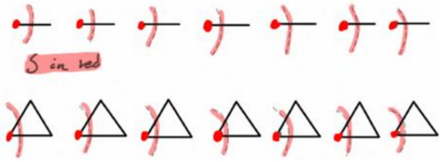
נגדיר את הגרף שהפונקציה מחזירה:

בהינתן φ נוסחת *3-CNF*, נגדיר גרף G_φ בדומה לרדוקציה של $\text{NAE-3-CNF-SAT} \leq_p 3COL$, אבל בלי הקודקוד D:



נגדיר את ה-k שהיא מחזירה:

נניח שיש ב-φ m פסוקיות ו-n משתנים. ראשית, אינטואיציה. אם נגדיר את S כדלהלן:

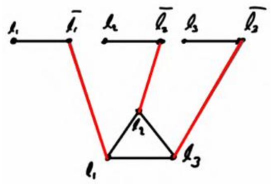


אז כל צלע בגאדג'ט של המשתנים תורמת צלע לחתך, וכל משולש תורם 2 צלעות לחתך. ובינתיים נקבל:

$$\sigma(G) = (\text{number of literals}) + 2(\text{number of clauses}) = n + 2m$$

נסתכל על הצלעות "באמצע" של G_φ, כלומר אלו שמחברות בין הגאדג'טים של המשתנים לגאדג'טים של הפסוקיות. כמה כאלו יש?

אם נתבונן בפסוקית שמיוצגת ע"י משולש, לכל קודקוד בה (שהוא ליטרל), יש שכן ייחודי (המשלים שלו) בגאדג'טים של הפסוקיות:



אז עבור m פסוקיות, נקבל $3m$ צלעות באמצע.

אנחנו לא יודעים איזה או כמה מהצלעות האלו חוצות את החתך. אבל בסה"כ, יש לכל היותר $3m+2m+n=5m+n$ צלעות בחתך.

נשים לב שאם נמצא $S\subseteq V(G_\varphi)$ כך ש $e_{G_\varphi}(S,\bar{S})\ge n+5m$, אז יתקיים:

$$e_{G_\varphi}(S,\bar{S})=n+5m$$

כי זה הגודל המקסימום של חתך אפשרי. כי כדי לקבל חתך גדול יותר, משולש יצטרך לתרום יותר משתי צלעות לחתך. וזה לא אפשרי. כלומר,

$$\sigma(G_\varphi)=e_{G_\varphi}(S,\bar{S})$$

נחזור להגדרת הרדוקציה: בהינתן φ, נגדיר $f(\varphi)=(G_\varphi,5m+n)$.

נבדוק שהפונקציה מקיימת את התנאים הנדרשים:

1 מתקיים באופן טריוויאלי מתהליך הבנייה.   3 מתקיים בדומה להסבר ברדוקציה של NAE-3-CNF-SAT $\le_p 3COL$.

נוכיח את 2 – נראה שמתקיים $(G_\varphi,5m+n)\in$ MAX-CUT $\Leftrightarrow \varphi\in$ NAE-3-CNF-SAT:

כיוון ראשון: נניח ש $(G_\varphi,5m+n)\in$ MAX-CUT, ויהי $(S,\bar{S})$ חתך המקיים $e_{G_\varphi}(S,\bar{S})=n+5m$. (נוכל לדרוש שוויון בעקבות המסקנה לעיל).

לפי אותה מסקנה, בגלל ש $e_{G_\varphi}(S,\bar{S})=n+5m$, נובע ש $\sigma(G_\varphi)=e_{G_\varphi}(S,\bar{S})$.

בפרט, חייב להתקיים עבור החתך שכל צלע בגאדג'ט של המשתנים היא צלע חוצה (קצה אחד ב-S והשני ב-$\bar{S}$).

נגדיר השמה a שתספק את φ ותהיה בתצורת $NAE\text{-}SAT$ כך:

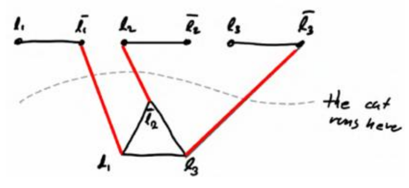
נתמקד בגאדג'טים של המשתנים. לכל צלע בהם נוודא ש-a מספקת את הליטרל שנמצא ב-S (ובהתאמה, לא מספק את הליטרל המשלים שנמצא ב-$\bar{S}$):

ההשמה תקינה (כל משתנה בעל ערך יחיד) כי קבענו את הערכים לפי הגאדג'טים של המשתנים וכל משתנה הופיע שם פעם אחת (עם המשלים).

נראה שהיא בתצורת $NAE\text{-}SAT$:

לכל פסוקית בגאדג'טים של הפסוקיות (המשולשים) יש 2 צלעות החוצות את החתך $(S,\bar{S})$.

בנוסף, כל צלע אמצעית ב-G_φ חוצה את החתך, ובפרט הם מהצורה $(\ell,\bar{\ell})$ - ליטרל והמשלים שלו.



אז לכל פסוקית יש ליטרל מ-S וליטרל מ-$\bar{S}$.

תהי C פסוקית כלשהי בגאדג'ט הפסוקיות, ויהי ℓ ליטרל המופיע בה.

אם $\ell\in S$, אזי הוא מחובר לליטרל $\bar{\ell}\in\bar{S}$ בגאדג'ט המשתנים, ולפי ההשמה a הוא יקבל את הערך T.

אם $\ell\in\bar{S}$, אזי הוא מחובר לליטרל $\bar{\ell}\in S$ בגאדג'ט המשתנים, ולפי ההשמה a הוא יקבל את הערך F.

מכיוון שכל פסוקית מכילה ליטרלים מ-S ומ-$\bar{S}$, אז בכל פסוקית יש ליטרל מסופק וליטרל לא מסופק. מש"ל.

כיוון שני: נניח ש φ בתצורת $NAE\text{-}SAT$ ותהי a השמת NAE שמספקת אותה. נמצא חתך ב-G_φ בגודל $5m+n$.

נגדיר את $S\subseteq V(G_\varphi)$ להיות קבוצת כל הקודקודים שהליטרלים שלהם מסופקים תחת a.

נשים לב ש-S מכילה קודקודים גם מהגאדג'טים של המשתנים וגם מהגאדג'טים של הפסוקיות.

נספור את הצלעות שחוצות את החתך:

מכיוון ש-a היא השמת NAE מספקת, בכל משולש בגאדג'ט הפסוקיות יש גם ליטרל שהוא T וגם ליטרל שהוא F. כלומר יש קודקוד מ-S וגם מ-$\bar{S}$.

לכן, כל פסוקית תורמת לפחות 2 צלעות החוצות את החתך.

בנוסף, מכיוון ש-a עקבית, כל צלע בגאדג'ט המשתנים חוצה את החתך (כי היא מוגדרת ע"י ליטרל והמשלים שלו).

לבסוף, כל צלע "אמצעית" ב-G_φ היא מהצורה $\left(\ell, \overline{\ell}\right)$ ולכן קודקוד אחד מסופק והשני לא, אז אחד ב-S והשני ב-$\bar{S}$. כלומר חוצה את החתך.

בסה"כ, יש לפחות $5m + n$ צלעות החוצות את החתך, ולפי הטענות שהוכחנו, יתקיים $\sigma\left(G_\varphi\right) \geq 5m + n$.

כלומר הוכחנו את התנאי השני של הפונקציה, אז הוכחנו ש $\text{MAX-CUT} \in NPH$ ולכן היא NPC.