אלגוריתמים מבוזרים - הרצאה 5, בעיית MIS

שרון מלטר, אתגר 17

22 בספטמבר 2024

תוכן עניינים

1 קבוצה בלתי תלויה מקסימלית (MIS)

קבוצה היא בת"ל מקסימלית אם היא בלתי תלויה (אף זוג צמתים אינם שכנים) ולכל קשת קיים צומת שבקבוצה (לכן היא לא חייבת להיות קבוצה עם הגודל הכי גדול)
למשל, בגרף כוכב, הצומת המרכזי מהווה קבוצה בת"ל מקסימלית. כמו כן גם קבוצה של כל הצמתים מסביבו היא קבוצה בת"ל מקסימלית.

זו בעיה מפורסמת, אשר מהווה דוגמה לאי־סימטריות ויש לה שימושים רבים, למשל בחירת מרכזים לקלאסטרים. נראה שימוש של MIS לצביעה.

1.1 אפליקציה לצביעה

1.1.1 למה

בהינתן אלגוריתם בלוקאל שמחשב MIS ולוקח $T(n)$ סיבובים עבור גרפים בעלי n צמתים (נקרא לו A), קיים אלגוריתם B בלוקאל שמחשב $(\Delta+1)$־צביעה עבור גרפים בעלי n צמתים ודרגה מקסימלית Δ ב־ $T(n\cdot(\Delta+1))$ סיבובים.

כמו כן נראה עוד מעט אלגוריתם רנדומי (עובד בהסתברות גבוהה) ב־ $O(logn)$ סיבובים כך שלפי הלמה שהצגנו קיים אלג' רנדומי ב־ $O(logn)$ סיבובים ל־ $(\Delta+1)$־צביעה בלוקאל (אך הוא מצריך הודעות גדולות)
בהרצאה הקודמת ראינו 2Δ־צביעה ב־ $CONGEST$ (הודעות קטנות) ב־ $O(logn)$ סיבובים.

1.1.2 הוכחת הלמה

אינטואיציה: יהיו G גרף קלט לצביעה. נסמלץ אלגוריתם ל־ MIS על גרף G' שמספר הצמתים בו הוא $n(\Delta+1)$ כך שנקבל ממנו $(\Delta+1)$־צביעה ל־ G.
פירוט: נבנה את G' כך־ נחליף כל צומת מ־ G נחליף בקליקה של $\Delta+1$ צמתים. כל קשת בין זוג צמתים ב־ G נחליף בקשת אחת מכל עותק של צומת לעותק המקביל של שכנו (במקום צומת יחיד ישנם $\Delta+1$ צמתים ובמקום קשת אחת יש $\Delta+1$ בסך הכל בין העותקים)
נחשב MIS עבור G'. נוכל לעשות זאת ב־ $T(n(\Delta+1))$ סיבובים, ניתן לסמלץ את האלגוריתם על G באופן הבא;

- לכל צומת v נסמלץ את הקליקה המתאימה ב־ G'.
- עבור הודעה שנשלחת על הקשת (u_i, v_i) נסמלץ ב־ G הודעה בקשת (u, v). ניתן לשלוח עד $\Delta+1$ הודעות על כל קשת. כיוון שגודל ההודעות לא חסום בלוקאל, ייקח סיבוב אחד לשלוח את כל ההודעות הרלוונטיות.
- הודעה שנשלחת על קשת מהצורה (v_i, v_J) ניתן לסמלץ לוקאלית עבור v כך שאין צורך בתקשורת.

כך נמיר את האלגוריתם שרץ על G' לאלגוריתם שרץ על G. מה נעשה עם אותו אלגוריתם?

נסמן ב־ S את ה־ MIS שנמצא. צומת v ייבחר בצבע i אם $v_i \in S$ (נוכיח שתמיד קיים בדיוק צומת אחד מהקליקה שב־ S)

הוכחת נכונות:
כל צומת בוחר צבע אחד לכל היותר (רק צומת אחד מהקליקה שלו נבחר ל־ S) מכיוון שלא קיים יותר מצומת אחד מקליקה ב־ S.
צמתים שכנים u, v בוחרים צבעים שונים מכיוון שאחרת גם u_i וגם v_i נמצאים ב־ S, אך הם שכנים.
כל צומת בוחר צבע, מכיוון שבכל קליקה יש $\Delta+1$ צמתים, והדרגה המקסימלית ב־ G היא Δ, כך שלא ייתכן שנבחרו יותר מ־ Δ שכנים של הקליקה של צומת u. ואז קיים בקליקה צומת שאף שכן שלו מהקליקה לא נבחר ל־ S וגם אף שכן שלו מקליקה אחרת לא נמצא ב־ S, כך ש־ S לא מקיימת את תכונת המקסימליות.
לכן מתקבלת $\Delta+1$־צביעה חוקית :)

כעת נעבור לאלגוריתם עצמו אשר מוצא MIS.

1.2 אלגוריתם למציאת MIS

נראה אלגוריתם רנדומי שלוקח $O(logn)$ סיבובים;

1. כל צומת בוחר מספר רנדומי $r_v \in [0,1]$ (המספר יכול להיות לא שלם, נרחיב על כך בקרוב) ושולח אותו לשכנים. הוא מצטרף לקבוצה S אם $r_u < r_v$ לכל שכן u של v.
את r_v ניתן לבחור מתחום סופי למשל $[1,n^2]$ כך שניתן לייצג אותו ב־ $logn$ ביטים. מטרתנו היחידה עבור מספר זה היא שבהסתברות גבוהה צמתים שונים יבחרו מספרים שונים.

2. אם v הצטרף ל־ S, הוא וכל שכניו עוזבים את האלגוריתם.
(מאותה נקודה הם כבר לא משתתפים באלגוריתם):)

הערה: אם במקום מספר רנדומי נבחר את r_v באופן דטרמיניסטי (שרירותי) למשל לפי הזהה, האלגוריתם עדיין יחשב MIS, אך הוא יכול להיות מאוד איטי לגרפים מסוימים. למשל, עבור גרף המסלול $(1,2,3,4,5,6,7)$ נצטרך מספר סיבובים לינארי, מכיוון שצמתים רבים אינם מקסימום מקומי בהתחלה (רק 7 בסיבוב הראשון) ואיך להם אפשרות להשתנות.
(נוכיח אחר כך את היתרונות ביעילים של רנדומיות)

וזהו. אלגוריתם חביב וחמוד :)

1.2.1 הוכחת נכונות

הקבוצה S הינה בלתי תלויה מכיוון שצומת v הצטרף ל־ S רק אם הוא מקסימום מקומי (מקסימום שגדול ממש משאר המספרים, כך שלא ייתכן ששני שכנים נכנסים ל־S באותו סיבוב), ולאחר מכן כל שכניו לא יכולים להיבחר. S מקסימלית מכיוון שצומת לא נמצא בקבוצה הינו שכן של צומת שהיה מקסימום מקומי ונכנס ל־ S. האלגוריתם מסתיים רק לאחר שכל הצמתים ב־ S או שכנים של צומת ב־ S.

אז עד כאן נכונות. מה לגבי יעילות?

1.2.2 משפט

האלגוריתם מסתיים תוך $O(logn)$ סיבובים בהסתברות של לפחות $1-\frac{1}{n}$.

1.2.3 הוכחה

<u>ניסיון 1:</u>
(הניתוח פה שגוי, אך נותן לנו אינטואיציה)
נרצה לחשב את מס' הצמתים שעוזבים את האלגוריתם בכל סיבוב. צומת u הוא מקסימום מקומי בהסתברות $\frac{1}{d(u)+1}$ (כאשר $d(u)$ היא הדרגה שלו) זאת מכיוון שבוחרים $r \in [0,1]$ כך שההסתברות לכך שלא כל המספרים של שכניו של u שונים זניחה. אם u הוא מקסימום מקומי, אז u וכל שכניו עוזבים את האלגוריתם ולכן

$$E[X] = \sum_{u \in V} \frac{d(u)+u}{d(u)+1} = \sum_{u \in V} 1 = n$$

כאשר X הוא מספר הצמתים שעוזבים.

הבעיה בניתוח: ייתכן שאנו סופרים את אותו הצומת מספר פעמים (עבור מספר שכנים שלו) לכן החישוב שגוי.

<u>ניסיון שני:</u>
במקום לנתח את הקטנת מספר הצמתים, ננתח את הקטנת מספר הקשתות.
נוכיח שבתוחלת מספר הקשתות קטן פי 2 בכל איטרציה, כך שנצטרך $O(logn)$ איטרציות (יש $O(n^2)$ קשתות)

נסמן ב־ G_i את הגרף בתחילת איטרציה (סיבוב) i וב־ m_i את מספר הקשתות בתחילת אותה איטרציה. נרצה להוכיח את הלמה הבאה;

$$E[m_{i+1} | G_i \ has \ m_i \ edges] \leqslant \frac{m_i}{2}$$

הקשת (u,v) עוזבת את האלגוריתם אם (u,v) או אחד השכנים שלהם u,v מצטרף ל־ S.
בשביל האנליזה, נחליף כל קשת לא מכוונת $\{u,v\}$ (סימון מקובל לקשתות לא מכוונות) בזוג קשתות $(u,v),\ \ (v,u)$.
כמו כן נרצה לשייך את העזיבה של קשת (u,v) לשכן w (של u) שהצטרף ל־ S עבורו המספר r_w מקסימלי כדי להימנע מספירות כפולות.
כלומר, באופן פורמלי, נגדיר את המאורע $A_{w\to u}$ בתור המקרה $r_w > r_v$ לכל $v \in N(w) \cup N(u)$.
w הוא גם מקסימום מקומי וגם מקסימלי בין השכנים של u.
אם $A_{w\to u}$ קרה, אז u,w (בין היתר) עוזבים את האלגוריתם ולכן גם כל ה־ $d(u)$ קשתות הסמוכות ל־ u. בסך הכל נקבל, נקבל שאם X הוא מספר הקשתות שעוזבות, אזי;

$$X_{w\to u} = \begin{cases} d(u) & if\ \ A_{w\to u}\ \ happens \\ 0 & else \end{cases}$$

$$X = \sum_{(w,u)\in E} X_{w\to u}$$

<u>טענה 1</u>: מספר הקשתות שעזבו הוא לפחות $\frac{X}{2}$.

הבחנה: אין ספירות כפולות, מכיוון שכל קשת נספרת ב־ X לכל היותר פעמיים (משתי הקשתות שמסמנות את שני הכיוונים שלה)

<u>הוכחה</u>: מהגדרת $A_{w\to u}$, אם $w \neq w'$ כאשר גם w' הוא שכן של u, לכל היותר אחד מהמאורעות $A_{w\to u},\ \ A_{w'\to u}$ יכול לקרות.
אם $A_{w\to u}$ קורה, אז כל הקשתות הסמוכות ל־ u עוזבות. כפי שצוין, הקשת $\{u,v\}$ יכולה להיות היותר להיספר פעמיים־ פעם אחת בגלל $A_{w\to u}$ ופעם שנייה בגלל $A_{w\to v}$ (קיים w יחיד מכיוון שהגדרנו אותו כמקסימום מקומי וכגדול ממספרי כל שכני u או v, תלוי במאורע)

<u>טענה 2</u>:

$$m_i \leqslant E[X]$$

<u>הוכחה</u>: $A_{w\to u}$ הוא המאורע בו $r_v < r_w$ לכל $v \in N(w) \cup N(u)$, כך שההסתברות אליו היא לכל הפחות $\frac{1}{d(w)+d(u)}$ (שוב, אנו מניחים כי ההסתברות לכך שלשני צמתים אותו מספר היא זניחה)
כך ש־

$$E[X] = \sum_{(w,u)\in E} E[X_{w\to u}]$$

$$E[X] \geqslant \sum_{(w,u)\in E} P(A_{w\to u} \cdot d(u)$$

$$E[X] \geqslant \sum_{(w,u)\in E} \frac{1}{d(u)+d(w)} \cdot d(u)$$

$$E[X] \geqslant \sum_{\{w,u\}\in E} \frac{d(u)}{d(w)+d(u)} + \frac{d(w)}{d(w)+d(u)}$$

$$E[X] \geqslant \sum_{\{u,v\}\in E} 1 = m_i$$

ומכאן נקבל ש־

$$E[edges\ \ that\ \ were\ \ removed] \geqslant \frac{E[X]}{2} \geqslant \frac{m_i}{2}$$

אזי בתוחלת מספר הקשתות קטן פי 2 בכל איטרציה. לכן תוחלת מספר האיטרציות היא $O(logn)$.
אבל רצינו גם להוכיח שזהו מספר הסיבובים בהתסברות **גבוהה**.

אחרי $4 \cdot logn$ איטרציות תוחלת הקשתות היא;

$$E[m_{4\cdot logn}] \leqslant \frac{m_0}{2^{4\cdot logn}} \leqslant \frac{n^2}{n^4} \leqslant \frac{1}{n^2}$$

ניעזר באי־שוויון מרקוב ונקבל;

$$Pr(M_{4\cdot logn} \geqslant 1) \leqslant \frac{1}{n^2}$$

ולכן בהתסברות גבוהה לא נשארו קשתות, והאלגוריתם מסתיים.

1.2.4 הערות

- באלג' של לובי צומת נכנס ל־ S אם הוא מקסימום מקומי. אלג' אחר $[Alon,\ Badani,\ Itay]$ באותה סיבוכיות מכניס צומת u ל־ S בהסתברות $\frac{1}{deg(u)}$.
- אינטואציה לאותו אלגוריתם:
נניח שלכולם אותה דרגה d, אז כל צומת מצטרף בהסתברות $\frac{1}{d}$. נצפה שבערך צומת אחד בסביבה יצטרף.