אלגוריתמים מבוזרים - הרצאה 8, תרגילים

שרון מלטר, אתגר 17

21 בספטמבר 2024

תוכן עניינים

1 MDS ע"י $Network\ Decomposition\ (ND)$ 3
1.1 תזכורת 3
1.2 פתרון 3
1.2.1 בניית ND עם התכונה הנוספת 4
2 שיפור הקירוב 5
2.1 עוד על ND 5
2.2 אלגוריתם סדרתי 5
2.3 אנליזת קירוב 5
2.4 אלגוריתם ב־ $LOCAL$ 6
2.5 סיבוכיות 6

נפתור מספר תרגילים הקשורים לנושאים שלמדנו :)

1 MDS ע"י $Network\ Decomposition\ (ND)$

1.1 תזכורת

ND היא חלוקת צמתים לקלאסטרים עם הזוג $(C,\ D)$ כך ש־ D הינו קוטר של קלאסטר ו־ C הוא מספר צבעים. קלאסטרים שכנים צבועים בצבעים שונים. (ניתן לחשוב על ND כעל הכללה של צביעה)

ND הינו כלי שימושי לפתרון בעיות לוקליות, כגון MIS. ראינו בהרצאה קודמת שניתן להשיג $C = O(logn),\ D = O(log^3n)$ בזמן $O(log^7n)$ סיבובים באלגוריתם דטרמיניסטי.
נרצה להשיג קירוב טוב ל־ MDS ע"י ND, כאשר MDS זוהי קבוצת צמתים בגודל מינימלי, כך שכל צומת נמצא בקבוצה או שכן של צומת מהקבוצה.

1.2 פתרון

ניסיון 1: נחשב פתרון אופטימלי לכל קלאסטר בנפרד וניקח את איחוד הפתרונות.
פתרון זה פיזבילי מכיוון שאם צמתי כל קלאסטר הם $dominated$ אזי כל הצמתים בגרף הם $dominated$.

הבעיה היא שלא מדובר בפתרון אופטימלי־ למשל, במקרה בו יש לנו גרף עם צומת שיש קשת ממנו לכל שאר הצמתים, הפתרון האופטימלי הוא קבוצה בגודל 1. אבל עם חלוקה לקלאסטרים, גודל הקבוצה שנקבל היא לפחות כמספר הקלאסטרים השונים.
נרצה למצוא תנאי כך שאם נמצא פתרון לקלאסטרים שונים במקביל, וניקח את איחוד הפתרונות, האיחוד יהיה לכל היותר בגודל $|OPT|$ (כאשר OPT זהו הפתרון האופטימלי לכל הגרף) שימו, לב, מתכוונים לאיחוד פתרונות של מספר קלאסטרים שונים, לא של **כל** הקלאסטרים. (אחרת הדרישה הייתה לא הגיונית מכיוון שפתרון לכל הגרף לא יכול להיות קטן מ־ OPT)
למשל, ב־ MIS יכולנו לפעול במקביל בקלאסטרים שהם לא שכנים כי הם לא משפיעים אחד על השני.

הבחנה: שני קלאסטרים לא משפיעים אחד על השני אם הם לא שכנים, ואם הם לא **חולקים** שכנים.
זאת מכיוון שבמקרה זה הוספת צומת מהם או משכניהם לקבוצה לא משפיעה על הצורך להוסיף או לא להוסיף צמתים בשביל הקלאסטר השני.

<u>טענה:</u> אם נחשב פתרון אופטימלי ל־ MDS עבור שני קלאסטרים במרחק 3 אחד מהשני, איחוד הפתרונות יהיה לכל היותר בגודל OPT.

<u>הוכחה:</u> יהא OPT פתרון אופטימלי לגרף כולו. נסמן ב־ OPT_1 את הפתרון האופטימלי לתת־גרף הכולל קלאסטר C_1 ואת שכניו וב־ OPT_2 את הפתרון האופטימלי עבור תת־גרף של קלאסטר C_2 ושכניו.
נוכיח ש־ $|OPT_1| + |OPT_2| \leqslant |OPT|$

הבחנה: הקבוצות $OPT_1,\ OPT_2$ הן זרות. זאת מכיוון ש־ $C_1,\ C_2$ אינם שכנים ולא חולקים שכנים. אזי $OPT_1,\ OPT_2$ הן תתי־קבוצות זרות של OPT, כך ש־ $|OPT_1| + |OPT_2| \leqslant |OPT|$ (נוכיח זאת פורמלית)

נסמן ב־ $OPT_1^*,\ OPT_2^*$ את הפתרונות האופטימליים עבור $C_1,\ C_2$ ושכניהם בהתאמה,
אז נובע $|OPT_1^*| \leqslant |OPT|,\ \ |OPT_2^*| \leqslant |OPT|$. זאת מכיוון שאם OPT_i הוא פתרון **כלשהו** לקלאסטר C_i ו־ OPT_i^* הוא פתרון **אופטימלי** לאותו קלאסטר, נקבל; $|OPT_1^*| + |OPT_2^*| \leqslant |OPT_1| + |OPT_2| \leqslant |OPT|$

ניתן להרחיב את הטענה באופן הבא:

<u>טענה:</u> אם יש קבוצת קלאסטרים $C_1,\ C_2, ...\ C_k$ כך שכל שני קלאסטרים הם במרחק לפחות 3 אחד מהשני ונחשב פתרון אופטימלי לכל קלאסטר בנפרד, איחוד הפתרונות יהיה לכל היותר בגודל $|OPT|$.

<u>רעיון הוכחה:</u> באותו אופן כמו בהוכחה הקודמת, נסמן ב־ OPT_i את צמתי OPT שהם ב־ C_i או שכניו. קבוצות

OPT_i שונות הן זרות, לכן נקבל $|OPT| \leqslant |OPT_i| \sum_{i=1}^{k}$.
בנוסף, לכל i יהי OPT_i^* פתרון אופטימלי ל־ c_i, אזי גודל הפתרון שחישבנו הוא $\sum_{i=1}^{k} |OPT_i^*| \leqslant \sum_{i=1}^{k} |OPT_i| \leqslant |OPT|$

הפתרון: לכן אם נחשב ND עם **התכונה הנוספת**, לפיה קלאסטרים מאותו צבע צריכים להיות במרחק של לפחות 3 אחד מהשני, ונחשב עבור אותם קלאסטרים פתרון אופטימלי במקביל, נקבל בסה"כ קירוב $O(logn)$.
מהטענה נקבל שלכל צבע בנפרד גודל איחוד הפתרונות הוא לכל היותר $|OPT|$, כך שהגענו לאן שרצינו :)
באלגוריתם לבניית ND שאנחנו מכירים ישנם $O(logn)$ צבעים, לכן נקבל $O(logn)$ קירוב לפתרון האופטימלי לכל הגרף.
נותר רק להראות כיצד נצבע את הגרף כך שקלאסטרים בצבע זהה יהיו במרחק של לפחות 3 אחד מהשני.

1.2.1 בניית ND עם התכונה הנוספת

כדי לקבל את התכונה הנדרשת, נחשב ND בגרף G^2; גרף עם אותם הצמתים של G, כך שיש קשת בין כל זוג צמתים שהמרחק ביניהם קטן־שווה ל־ 2. נקבל שכל שני קלאסטרים שהינם שכנים ב־ G^2 יהיו בצבעים שונים.

נניח ש־ u, v הם שני צמתים ממרחק הקטן־שווה ל־2, אזי יש ביניהם קשת ב־ G^2, כך ש־ u, v באותו קלאסטר או שהם בקלאסטרים בצבעים שונים. אם יש שני צמתים שהם בקלאסטרים **שונים** ו **באותו צבע** אזי הם במרחק של לכל הפחות 3 אחד מהשני. לכן כל הקלאסטרים מאותו צבע הם ממרחק לכל הפחות 3 אחד מהשני, כנדרש.

ניתן לחשב פתרונות אופטימליים עבור קלאסטרים מאותו הצבע במקביל ולקבל שאיחוד הפתרונות הינו לכל היותר $|OPT|$, כפי שרצינו :)

אך איך נריץ אלגוריתם ND ב־ G^2?
נסמלץ אותו בלוקל על הגרף הקיים, G, בתוספת של $O(1)$ סיבובים לכל סיבוב של בניית ND ב־ G.
"נגדיר מחדש" את שכניו של צומת כצמתים שבמרחק לכל היותר 2 ממנו. אזי נצטרך שני סיבובים כדי לתקשר עם כל שכניו של צומת, כלומר נכפיל את מספר הסיבובים, וסיבוכיות בניית ה־ ND היא עדיין $O(log^7 n)$.

אז קיבלנו את כל מה שרצינו. זה הזמן לרצות יותר.
נשפר את הקירוב שקיבלנו.

2 שיפור הקירוב

נרצה להראות קירוב $1+\epsilon$.

2.1 עוד על ND

ND הינו כלי שימושי לסימלוץ אלגוריתמים סדרתיים חמדניים עם לוקליות 1 (נדבר בקרוב על מה זה אומר) לדוגמה MIS ו־ $(\Delta+1)$־צביעה. במקרים אלו קיים אלגוריתם סדרתי שפשוט עובר על הצמתים אחד אחרי השני ומחליט על הפלט בהתאם לשכנים (סביבת 1 :))
אפשר להרחיב אותו לאלגוריתמים סדרתיים עם לוקליות r (תלות בסביבת r) אם במקום לחשב ND בגרף G נחשב ND בגרף G^r (ואז כל סביבת r של צומת כוללת את כל שכניו) במודל ה־ $LOCAL$ זה יוסיף פקטור r לסיבוכיות (נצטרך r סיבובים כדי לתקשר עם שכניו של צומת) נסמלץ סיבוב ב־ G^r ע"י $O(r)$ סיבובים ב־ G.
ניעזר ברעיון הזה כדי להראות אלג' $(1+C)$־קירוב ל־ MDS. נתחיל בתיאור **אלגוריתם חמדן סדרתי** עם לוקליות r עבור $r=O(\frac{logn}{\epsilon})$.

הערה: הקירוב הרבה יותר טוב ממה שאלג' סדרתיים פולינומיים יכולים להשיג.

2.2 אלגוריתם סדרתי

נתחיל מכמה הגדרות שניעזר בהן;

- יהא $v_1,\ v_2,...,\ v_n$ סדר על הצמתים.
- $B_r(v)$ ־ זהו הכדור ברדיוס r סביב v.
- $g(v,\ r)$ ־ זהו גודל של MDS אופטימלי עבור כל הצמתים **שלא מכוסים** ב־ $B_r(v)$.

במהלך האלגוריתם נבנה MDS המסומן D וחלק מהצמתים כבר מכוסים מאיטרציות קודמות.

הערה: הפתרון האופטימלי הנ"ל יכול להשתמש גם בצמתים מחוץ ל־ $B_r(v)$, אך הם חייבים להיות מ־ $B_{r+1}(v)$ מהגדרת בעיית MDS.

וכעת הרגע שכולנו חיכינו לו, האלגוריתם הסדרתי;

1. עוברים על הצמתים לפי הסדר.
2. באיטרציה i, אם v_i עדיין **לא** מכוסה, מוצאים r_i כך ש־ $g(v_i,\ r_i+2)\leqslant(1+\epsilon)\cdot g(v_i,\ r_i)$. כמו בהרצאה הקודמת, ניתן להראות ש־ $r_i=O(\frac{logn}{\epsilon})$ (הגבלה על מס' הפעמים שניתן להגדיל את הסביבה) רעיון: סדרת הערכים $g(v_i,\ r_i)$ מונוטונית עולה עם r_i וחסומה ע"י n, הערכים לא יכולים לגדול "יותר מדי" מ־ $logn$ פעמים.
3. אחרי שמצאנו r_i העומד בתנאים, מוסיפים את ה־ MDS האופטימלי של $B_{r_i+2}(v_i)$ ומסמנים צמתים אלו **כמכוסים** ־ לא נעבור עליהם שוב כחלק מהאלגוריתם, אבל כן ניתן להשתמש בהם כדי לכסות את שכניהם.

ניתן לראות שהפתרון פיזבילי. אך מה לגבי הקירוב?

2.3 אנליזת קירוב

נסמן: D^* הינו הפתרון האופטימלי. V_i היא קבוצת הצמתים מ־ $B_{r_i}(v_i)$ שאינם מכוסים בתחילת האיטרציה ה־ i. $D_i^*\subseteq D^*$ היא תת־קבוצה מינימלית מ־ D^* שמכסה את V_i.

טענה: הקבוצות D_i^* זרות.

הוכחה: מהגדרת האלגוריתם צמתים שהם בקבוצות $V_i,\ V_j$ כך ש־ $i\neq j$ הם במרחק של לכל הפחות 3 זה

מזה.
למה זה טוב? ־ לצמתים שמרחקם זה מזה הוא לכל הפחות 3 אין שכנים משותפים, אזי הקבוצות $D_i^*, \ D_j^*$ זרות כנדרש :)

ולמה הטענה הזו טובה? ־ להוכחת הקירוב!
מהגדרת האלגוריתם, באיטרציה i הפתרון שהוספנו הוא לכל היותר מגודל $(1+\epsilon)\cdot|D_i^*|$. אם נסכום את כל האיטרציות, מאחר שהקבוצות $D_i^*, \ D_j^*$ זרות, נקבל;

$$\sum_{i=1}^{n}(1+\epsilon)\cdot|D_i^*| \leqslant (1+\epsilon)\cdot|D^*|$$

איזה אושר. אבל!
עד כאן דיברנו וניתחנו אלגוריתם **סדרתי**, מה לגבי אלגוריתם בלוקל כפי שהובטח?

2.4 אלגוריתם ב־ $LOCAL$

כדי לקבל אלגוריתם ב־ $LOCAL$ נשתמש ב־ ND. נבחר $r = O(\frac{logn}{\epsilon})$ כך ש־ $2\cdot(r_i+3) < r$ לכל r_i (r הוא חסם שניתן לחשב מתוך n ו־ ϵ, אין צורך בידיעת r_i)

נגדיר גרף G^r, כך שצמתיו הם צמתי G וקיימת קשת בין כל צמתים שהמרחק ביניהם ב־ G הוא לכל היותר r.
ניתן לסמלץ אלגוריתם ב־ $LOCAL$ על G^r בעזרת הרצת $O(r)$ סיבובים על G לכל תקשורת עם שכנים ב־ G^r.

ניתן לצומת תווית $(col_v, \ id_v)$ (col_v הינו הצבע של v ב־ ND) ונגדיר סדר לצמתים.

נסמלץ את האלג' **הסדרתי** שלמדנו בהתאם לסדר שהגדרנו לפי סדר עולה של צבעים:
באיטרציה j נפעל במקביל בכל הקלאסטרים מצבע j. נסמלץ את האלג' הסדרתי בכל קלאסטר כזה ע"י לימוד הקלאסטר ושכניו. בסוף החישוב כל צומת בעל צבע j אומר לשכניו ב־ G^r איזה צמתים נוספו ל־ MDS.
נראה שבאמת ניתן לסמלץ את האלגוריתם הסדרתי עבור כל הצמתים בעלי צבע j במקביל־

- נסתכל על צומת v_i בעל צבע j: החישוב של v_i באלג' תלוי רק בסביבת r_i+3 שלו (ובחרנו $r_i+3 \leqslant r$)
- מוצאים r_i כך ש־ $g(v_i, \ r_i+2) \leqslant (1+\epsilon)\cdot g(v_i, \ r_i)$
(כזכור, ניתן להראות ש־ $r_i = O(\frac{logn}{\epsilon})$)
- כל הצמתים שהוא תלוי בהם הם **שכנים** שלו ב־ G^r. אם הם **בצבע שונה** הם יטופלו באיטרציה אחרת והוא כבר יודע את הפלט של כל אלה שלפניו כנ"ל מאיטרציות קודמות.
אם הם **באותו צבע** אזי הם חייבים להיות באותו הקלאסטר מהגדרת ND ב־ G^r. בכל קלאסטר אם מסמלצים את האלג' הסדרתי לפי הסדר שם.

בסה"כ קיבלנו שהאלג' מסמלץ את האלגוריתם הסדררתי ולכן ישיג $(1+\epsilon)$ קירוב.

2.5 סיבוכיות

ב־ G^r יש C איטרציות (אחת לכל צבע) שלוקחות $O(D)$, קוטר הקלאסטרים ב־ G^r. סמלוץ כל סיבוב ב־ G^r מצריך $O(r)$ סיבובים, ולכן בסה"כ סיבוכיות הסיבובים היא $O(rDC)$, כלומר $O(\frac{poly(logn)}{\epsilon})$