שימושי אלגברה לינארית במדעי המחשב

מבחן בית

בבחינה זו נתמקד בבעיות חישוביות המשלבות גרפים, והילוכים מקריים.
הבחינה תשלב שאלות תיאורטיות ושאלות מעשיות אותן תצטרכו לפתור תוך כדי שימוש בתכנית מחשב שאותה תממשו.

לאורך כל סעיפי הבחינה לכל גרף $G = (V, E)$, נסמן:
$V = \{1, \dots, n\}$ ו- $E \subseteq V \times V$
כלומר, תיתכנה קשתות עצמיות וקשת תיוצג על ידי זוג סדור $(i, j) \in V \times V$

בבחינה נחקור מספר פרמטרים של אלגוריתם ה- *PageRank* המתואר להלן בקצרה.

אלגוריתם ה- *PageRank*:

קלט: גרף מכוון G.

פלט: וקטור הסתברות $\vec{v} \in \mathbb{R}^n$ המייצג דירוג של קודקודי הגרף.

פרמטרים נוספים: N, t, p.

מבנה נתונים: מערך d בכגודל n המאותחל לאפסים.

תיאור האלגוריתם:
האלגוריתם יריץ t איטרציות, כאשר בכל איטרציה נבצע:

> נבצע הילוך מקרי באורך N צעדים, כאשר בכל צעד בהילוך:
> בסיכוי $1 - p$ נלך לשכן מקרי של הקודקוד בו אנחנו מבקרים (אם אין לו שכנים נשאר בקודקוד זה בסיכוי $1 - p$) ובסיכוי p נלך לקודקוד מקרי.
> יהי i הקודקוד אליו הגענו לאחר N צעדים, נעלה את ערכו של $d[i]$ ב- 1.

בתום ריצת האלגוריתם, נחלק את ערכי d ב- t ונחזיר וקטור שערכיו הם ערכי המערך d.

מתי נעצור את האלגוריתם?
ננסה לחקור זאת בעבודה זו.

בבחינה, נבחן את פלטי האלגוריתמים על משפחות גרפים שונות ונבחן את השפעתם של הפרמטרים N, t, p על קלטים אלו ואת הקשר שלהם לתנאי העצירה של האלגוריתם.

שאלה 1:

נסתכל על איטרציה בודדת (הילוך מקרי באורך N צעדים):

א. מה תוחלת מספר הצעדים בהם נבחר ללכת לקודקוד מקרי בגרף (ולא לשכן מקרי)?

ב. הסבירו מדוע אם נבחר $N \ll 1/p$, הפרמטר p חסר חשיבות מעשית.

ג. הסבירו מדוע התפלגויות הקודקדים לאחר הילוך באורך $1/p$ צעדים ולאחר $2/p$ צעדים תהינה דומות (הדרכה: מה קורה להתפלגות הקודקודים לאחר שהלכנו לקודקוד מקרי ולא לשכן מקרי).

ד. באיזה מבני נתונים תשמשו על מנת לייצג את הגרף על מנת לייעל את ריצת האלגוריתם? כתבו חסם אסימפטוטי הדוק ככל הניתן לזמן הריצה של איטרציה בודדת של האלגוריתם?

משפחות הגרפים בדיון:

משפחה ראשונה בדיון – גרפים מקריים:

בחלק זה נסתכל על גרפים המוגדרים מתוך ההתפלגות הבאה.

נקבע $0 < q < 1$ ולכל זוג קודקודים: (i, j) בגרף, נוסיף את הקשת (i, j) בסיכוי q.

משפחה שניה בדיון – מודל אחר לגרפים מקריים:

בחלק זה נסתכל על גרפים המוגדרים מתוך ההתפלגות הבאה.

לכל קודקוד i תותאם הסתברות q_i.

לכל זוג קודקודים (i, j) בגרף נוסיף את הקשת (i, j) בהסתברות q_j.
(כלומר, ההסתברות להוספת הקשת (i, j) תלויה בהסתברות המותאמת לקודקוד הנכנס j).

שאלה 2:

עבור כל אחת ממשפחות הגרפים שתוארו:
כיצד אתם מצפים שתיראה התפלגות הקודקודים בהילוך מקרי? כיצד באה לידי ביטוי התלות בפרמטרים של ההתפלגויות ובפרמטרים של האלגוריתם?

תנאי העצירה של האלגוריתם:

נריץ את האלגוריתם שלנו באופן הבא:

1. נריץ אותו עם פרמטר: $t = 2$.

2. נריץ אותו עם פרמטר: $t = 4$.

3. נריץ אותו עם פרמטר: $t = 8$.

...

מתי נעצור את הריצה? כאשר המרחק בין הרצות עוקבות היה קטן.

<u>תנאי עצירה ראשון</u> (עם פרמטר ϵ):

יהי v הווקטור שהתקבל ברצה עם פרמטר $t = 2^i$
ו- v' הווקטור שהתקבל עבור $t = 2^{i-1}$ אם:

$\|v - v'\| < \epsilon$ נעצור.

<u>תנאי עצירה שני</u> (עם פרמטר k):

נסתכל על k הקודקודים שדורגו ראשונים בהרצה עם פרמטר $t = 2^i$
ועל k הקודקודים שדורגו ראשונים בהרצה עם פרמטר $t = 2^{i+1}$, אם
קבוצות הקודקודים זהות ודירוגם זהה נעצור.

(כלומר, אנחנו מתעניינים רק ב- k הקודקודים שדורגו ראשונים ואם הדירוג נותר יציב, נעצור).

<u>שאלה 3:</u>

הגרילו 3 גרפים ממודל ההתפלגות הראשון עם הפרמטרים הבאים:
מספר הקודקודים בגרף: 2^{12}.
ההסתברות q לבחירת קשתות בגרף הראשון הינה $1/2^{12}$, בשני $1/2^9$ ובשלישי $1/2^4$.
הריצו את האלגוריתם עם ערכי $p = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^5}$, וכן: $N = 1/p$
הריצו פעם אחת עם תנאי העצירה הראשון עבור $\epsilon = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^5}$
ופעם עם השני עבור $k = 2,4,8, \ldots, 2^5$.

א. מה היה ערכו של t בכל אחת מההרצות?

ב. נסו לקשר בין תנאי העצירה השונים: עבור כל ערך של ϵ מה היה מספר הקודקודים שנותרו בדירוג יציב כשהאלגוריתם נעצר ולהיפך.

ג. תארו את התפלגות הקודקודים שהתקבלה? אפיינו את קבוצת הקודקודים שדורגו הכי גבוה.

ד. מדוע והאם תוצאת החישוב השתנתה כפונקציה של הפרמטר p בכל אחת מההרצות?

<u>שאלה 4:</u>

הגרילו 2 גרפים ממודל ההתפלגות השני.
מספר הקודקודים בכל אחד מהגרפים: 2^{12}.
בגרף הראשון לקודקוד i תותאם ההסתברות $q_i = \frac{1}{i}$.
בגרף השני לקודקוד i תותאם ההסתברות: $q_i = \frac{1}{\sqrt{i}}$

הריצו את האלגוריתם עם ערכי $p = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^5}$, וכן: $N = 1/p$
הריצו פעם אחת עם תנאי העצירה הראשון עבור $\epsilon = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{5}$
ופעם עם השני עבור $k = 2,4,8, \ldots, 2^5$.

חזרו על הסעיפים א'-ד', מהשאלה הקודמת.

<u>הערות כלליות:</u>

1. חשבו כיצד יש להציג את תוצאות הניסויים שערכתם בצורה בהירה וקריאה.
2. בכל אחת מהשאלות נסו להסביר את השוני (או הדמיון) שהתקבלו בתוצאות עם פרמטרים שונים.
3. אתן מוזמנות לבצע ניסויים נוספים, להגדיר משפחות נוספות של גרפים תוך כדי תיאור מדויק של הניסויים שביצעתן.

בהצלחה! ☺