

מעבדה 6 - הנחיות הגשה

עליכם להגיש סיכום של המעבדה **במסמך PDF או HTML אחד** כתוצר של מחברת ג'ופיטר - ניתן יהיה להגיש רק קובץ אחד.
המחברת תכלול קוד, טבלאות רלוונטיות, ויזואליזציות, את הסיכומים והתשובות ניתן לכתוב בכתב יד ולהוסיף כתמונות למסמך (בתנאי וקריא) או להקלידן. יש לתעד את הקוד באופן ברור עבור כל פעולה.
יש לתת **כותרת ברורה וקריאה עבור כל גרף וטבלה** (ולא כותרת כללית לכלל הגרפים מאותו סוג) מקרא עבור הגרפים, כותרות לצירים (מומלץ גם משפט הסבר).
במידה והתבקשתם להציג ערכים, בבקשה לכתוב במשפט את המסקנה מהגרף.
מומלץ לבנות את המחברת בצורה כללית כך שתהיה נכונה לכל אוסף מסמכים (ותהפוך להיות כלי עבודה עבור כל ניתוח טקסטואלי)

כנתונים לניתוח עליכם להשתמש בנתונים סינטטיים היוצרים **רשת קשרים באופן רנדומלי ועל 20 צמתים לפחות**. בניית הרשת כפי שהוצג בשיעור במחברת `simple pageRank.ipynb`

1. הסבירו וסכמו על LinkAnalysis במילים שלכם, ובהתייחס לנלמד בהרצאה ובמסלול, באופן כללי, וספציפית לטיפול בטקסט.
2. סכמו והסבירו את סוגי האלגוריתמים השונים שנלמדו בהרצאה (לכל הפחות PageRank ו HITS)
 1. מתי לדעתכם נשתמש בכל אחד ? (למה הם מיועדים, ודוגמה לתרחיש שימוש)
 2. מהי צורת הייצוג לנתונים ?
 3. לאיזה תהליכי עיבוד מקדים נדרש ?
 4. יתרונות, חסרונות, סקלביליות זמני ריצה וכו'.
3. בעבור אחת מהחבילות לבחירתכם [/https://networkx.org](https://networkx.org) ו [/https://igraph.org/python](https://igraph.org/python)
 1. קיראו וסכמו על צורות הייצוג השונות שיש לנו לנתונים (סוגי גרפים). יש לפרט באופן כללי, ולהרחיב בצורות הייצוג הרלוונטיות לנו, כפי שנלמד בהרצאה.
סכמו על התכונות, שיטות, פרמטרים, אופי הפעולה, וההפעלה ברמה הטכנית של כל אחת מצורות הייצוג הנ"ל.
 2. קיראו וסכמו על אלגוריתמי ה LinkAnalysis השונים. (PageRank ו HITS) סכמו על התכונות, שיטות, פרמטרים, אופי הפעולה, וההפעלה ברמה הטכנית של כל אחד מהאלגוריתמים.
 3. קיראו וסכמו על אפשרויות הוויזואליזציה שלנו, כולל דרכים להוסיף כותרות והסברים לצמתים וקשתות. יש להרחיב ולפרט את האפשרויות השונות.
4. חקר החבילה שנבחרה בסעיף קודם ברמה המעשית:
 1. בחלק זה עליכם להדגים את אופי העבודה של כל אחת מהחבילות ומציאת צמתים מרכזיים בעזרת PageRank ו HITS
 2. יש להציג זמני בניית צורות הייצוג, ועבודת האלגו'.
 3. יש להציג את הגדלים השונים של הגרפים שנוצרו.
 4. יש להציג וויזואליזציות רלוונטיות של הגרפים.
5. סכמו את התוצאות באופן השוואתי וברור והסבירו במילים שלכם (יש להתייחס לנקודות הבאות):
 - i. ניתוח של אופי ייצוג הגרף.
 - ii. עבודה עם האלגוריתמים השונים.
 - iii. בדיקת זמני ריצה וגודל מבני נתונים רלוונטיים.
 - iv. מסקנות

6. חלק ג' - סיכום

1. עליכם לכתוב סיכום קצר במילים שלכם (4-5 פסקאות), מתוך התייחסות לידע האישי שלכם בתחום מדעי הנתונים, והחומר הנלמד עד היום בתואר.
2. מה למדתם ממעבדה זו - יש להציג טבלאות השוואתיות וגרפים (גדלים ומדידות זמנים).
3. יש לדון בתוצאות הניסויים השונים שלכם.
4. יש לסכם את המסקנות הנובעות מההשוואות השונות.
5. עליכם להסביר מעבדה זו בהתייחס לחומר התיאורטי שנלמד עד היום בתחום מדעי הנתונים.

7. בונס עד 20 נק, ניתן לקבל עד 120 נק בעבודה זו ולהשפיע על **ציון העבודות הסופי**:
חלק זה יש לממש בעזרת החבילה **שלא נחקרה** בסעיפים 3-4 .
הציגו בעזרת הנתונים להלן את גרף הקשרים המשפחתיים וקשרי העבודה של משפחות האצולה של פירנצה מתקופת הרנסנס, כפי שנחקרו על ידי פדגייט וחבריו.
נתחו את הנתונים והציגו את המשפחות המרכזיות בכל תחום, והביעו את מסקנותיכם.

This example introduces the PageRank centrality measure with a network of Renaissance Florentine families around 1430. Specifically, we examine the PageRank centrality of the Florentine families in their marriage network. The families are nodes, and marriage ties between the families are edges in the network

http://www.casos.cs.cmu.edu/computational_tools/datasets/external/padgett/index2.html

בהצלחה !