

שם הקורס: למידה מתוך גרפים ומודלים דלילים

פרטי מרצה
שם: וסים חליחל
חדר: 115
טלפון: 0545975670
דואר אלקטרוני: wasimh@tauex.tau.ac.il

תיאור כללי של הקורס

בעיית חיפוש מבנים חבויים במידע ממימד גבוה, בו מימד המידע גדול ממספר הדגימות הזמינות, צברה תאוצה בשנים האחרונות. מטרת הקורס היא הקנייה של רקע מתמטי לניתוח בעיות מסוג זה, בעיקר במקרים בהם המידע מיוצג על ידי מטריצה ו/או גרף. בחלק הראשון של הקורס, תוך שימוש בשיטות מתחום תורת-המידע, נפתח חסמים סטטיסטיים, ואלגוריתמים המשיגים את אותם החסמים (תוך התעלמות מהביטים הישויים), עבור בעיה נתונה עם מבנה חבוי. למרות שבעיות ממימד גבוה עם מבנה מאפשרות פתרון סטטיסטי, האלגוריתמים המשיגים פתרונות אלו הם בעלי סיבוכיות חישובית גבוהה מאוד. מצדו השני של המתנס, אלגוריתמים בעלי סיבוכיות חישובית נמוכה דורשים יותר דגימות באופן משמעותי. בחלק השני של הקורס, נפתח וננתח אלגוריתמים יעילים לאותן בעיות, וכן נסביר את הפערים בין הגבולות הסטטיסטיים לגבולות החישוביים על ידי מיפויים מבעיות קשות מתורת הסיבוכיות. הקורס יכסה חלק מהנושאים הבאים: בעיית הקליקה החבויה, שערך וקטור דליל ממדידות לינאריות, בעיית גילוי הקהילות, בעיית שערך תת מטריצה, בעיית ניתוח גורמים ראשיים דלילים, ושערך מטריצה מדרגה נמוכה.

נושאי הקורס

שבוע	נושא
1.	סקירה כללית (נושאי הקורס וכלים)
2.	חזרה על תורת ההסתברות ואי שוויונים של ריכוז מידות
3.	הגדרה של בעיית הגילוי/שערך, ובעיית הקליקה (הצגת הבעיה ותכונות)
4.	בעיית מציאת מקסימום קליקה והקליקה החבויה (הצגה וניתוח של אלגוריתמים סטטיסטיים ויעילים)
5.	חזרה על יסודות תורת המטריצות האקראיות
6.	חזרה על יסודות השיטה הספקטרלית (תורת הפרטובציות)
7.	הגישה הספקטרלית ו SDP לבעיית הקליקה החבויה
8.	מודל גרף סמי-רנדומי והשערת הקליקה החבויה
9.	בעיית גילוי הקהילות (אלגוריתמים ומעברי פאזה)
10.	בעיית גילוי הקהילות (המשך) ובעיית שערך תת מטריצה חבויה
11.	דיון על הפערים בין הגבולות הסטטיסטיים לחישוביים
12.	בעיית ניתוח גורמים ראשיים דלילים
13.	שערך וקטור דליל ממדידות לינאריות

קורסי קדם

תורת ההסתברות, אותות אקראיים ורעש.

מדיניות ודרישות הקורס

הציון הסופי יקבע על סמך נוכחות, מטלות בית, וכן עבודת גמר וסמינר בהתבסס על הנושאים הנלמדים. בקורס.

מרכיבי הציון

מרכיב	אחוז מהציון
נוכחות	30%
מטלות בית	30%
עבודת גמר וסמינר	40%
סה"כ	100%

Course Title: Learning in Graphs and Sparse Models

Lecturer: Wasim Huleihel
Office: 115
Phone number: 0545975670
Email: wasimh@tauex.tau.ac.il

Course Description

The problem of searching for latent structures in high-dimensional data, where the dimensionality of the data is at least as large as the sample size, received a growing attention in the last few years. The primary goal of this course is to give a mathematical background for studying such problems, focusing on data that may be represented by matrices and/or graphs. To that end, in the first part of this course, we will discuss information-theoretic methods for establishing the fundamental/statistical limits (i.e., putting computational considerations aside) for a given inference problem with latent structure. Although structural assumptions in high-dimensional problems can yield nontrivial estimation rates, the statistically optimal estimators for these problems are typically not efficiently computable. Conversely, all known efficient algorithms for these problems are statistically suboptimal, requiring more data than strictly necessary. Accordingly, in the second part of this course, we will develop computationally efficient algorithms that in some cases attain the statistical limits, while in other cases cannot. We will discuss a recent line of work that aim to explain these gaps between what can be achieved statistically and computationally, through reductions from conjecturally hard problems in complexity theory. The course will cover a subset of the following topics, time permitting: planted clique and partition problem, estimating a sparse vector from linear observations, community detection, subgraph and matrix recovery problems, sparse principal component analysis, and low-rank matrix estimation.

Tentative Outline

Week	Topic
1.	Overview (topics and techniques covered in this course)
2.	Probability theory and concentration inequalities
3.	Formulation of the recovery/detection problem, and the clique problem (introduction & max clique)
4.	The maximum clique problem & planted clique problem (statistical and efficient algorithms)
5.	Introduction to random matrix theory
6.	Introduction to the spectral method (perturbation theory)
7.	Spectral method & SDP for planted clique
8.	The semi-random model & planted clique conjecture
9.	The community detection problem (algorithms and phase transitions)
10.	The community detection problem (cont'd) and biclustering
11.	Statistical-computational gaps
12.	Sparse principle component analysis
13.	Sparse principle component analysis

Course Prerequisites

Probability theory, Random signals and noise.

Grading

Item	Grade
Participation	30%
Homeworks	30%
Final project and seminar	40%

ספרות הקורס וביבליוגרפיה (Bibliography)

- Yihong Wu and Jiaming Xu, “Statistical Inference on Graphs: Selected Topics”, Lecture notes created for S&DS 684, 2019.
- Noga Alon, Michael Krivelevich, and Benny Sudakov, “Finding a large hidden clique in a random graph”, *Random Structures & Algorithms*, 13(3-4), 457-466, 1998.
- Bela Bollobas, “Random graphs”. Cambridge university press, 2001.
- Q. Berthet and P. Rigollet, “Complexity theoretic lower bounds for sparse principal component detection”, *Journal of Machine Learning Research: Workshop and Conference Proceedings*, 30, 1046-1066, 2013.
- Yudong Chen and Jiaming Xu, “Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices”, *The Journal of Machine Learning Research*, 17(1), 882-938, 2016.
- Yael Dekel, Ori Gurel-Gurevich, and Yuval Peres, “Finding hidden cliques in linear time with high probability”, In *Proceedings of the Meeting on Analytic Algorithmics and Combinatorics, ANALCO '11*, pages 67-75, Philadelphia, PA, USA, 2011. Society for Industrial and Applied Mathematics.
- U. Feige and R. Krauthgamer, “Finding and certifying a large hidden clique in a semirandom graph”, *Random Structures & Algorithms*, 16(2), 195-208, 2000.
- B. Hajek, Y. Wu, and J. Xu, “Information limits for recovering a hidden community”, *IEEE Trans. on Information Theory*, 63(8), 4729-4745, 2017.
- B. Hajek, Y. Wu, and J. Xu, “Achieving exact cluster recovery threshold via semidefinite programming” *IEEE Transactions on Information Theory*, 62(5), 2788-2797, May 2016.
- Michael Krivelevich and Benny Sudakov, “The largest eigenvalue of sparse random graphs”, *Combinatorics, Probability and Computing*, 12(1), 61-72, 2003.
- Elchanan Mossel, Joe Neeman, and Allan Sly, “Reconstruction and estimation in the planted partition model”, *Probability Theory and Related Fields*, 162(3-4), 431-461, 2015.
- Matthew Brennan, Guy Bresler, and Wasim Huleihel, “Reducibility and Computational Lower Bounds for Problems with Planted Sparse Structure”, *Proceedings of the 31st Conference On Learning Theory*, 48-166, 2018.
- Matthew Brennan, Guy Bresler, and Wasim Huleihel, “Universality of Computational Lower Bounds for Submatrix Detection”, *Proceedings of the 32st Conference On Learning Theory*, 417-468, 2019.
- Jerry Li, “Robust sparse estimation tasks in high dimensions”, *Proceedings of the 30st Conference On Learning Theory*, 2017.
- Sivaraman Balakrishnan, Simon S Du, Jerry Li, and Aarti Singh, “Computationally efficient robust sparse estimation in high dimensions”, *Proceedings of the 30st Conference On Learning Theory*, 2017.