Metric Embedding and Sketching

COSC 249 (Fall 2022)

Hsien-Chih Chang
HSIEN-CHIH.CHANG@DARTMOUTH.EDU

Time and Place

Tuesday Thursday 2:25pm – 4:15pm (2A)
(location to be updated)

Course Description

In data analysis we can often assume the input is drawn from a *metric space* associated with some well-behaved distance function. In such scenario one can hope to find an alternative representation — an *embedding* — of the input data without sacrificing the distance information too much. To our surprise, not only this is possible, but often times one can also perform a *sketching* to reduce the size and amount of the data required.

This graduate-level seminar-style course is aimed to introduce the various ways to encode metric spaces in a succinct fashion with minimal distortion, suitable for their algorithmic purposes. Naturally, due to the vast amount of work and literature in the area, the topics covered in this class will be biased towards the interest and expertise of the instructor.

Course Objectives

The course serves as an opportunity for people interested in the theory of metric embeddings and their algorithmic applications to learn about the results and techniques in recent years.

Prerequisite

Students are assumed to have the math maturity to read and write proofs and establish basic properties of newly encountered mathematical structures. *COSC 30 Discrete Math*, *COSC 31 Algorithms*, and at least one course in probability (say Math 20 or COSC 49.10) are required as prerequisite. Previous exposure in approximation algorithms (COSC 36) or graph theory (MATH 38) will be very helpful. We will hand out homework 0 in the first week and give the students an idea of the background knowledge expected.

Textbook

The following notes will be our main reference for the first part of the class:

- *Lecture notes on metric embeddings* by Jiří Matoušek.

The second part of the class will be on various topics in metric sketching, based on recent research in the field. Relevant reading materials will be linked on the course webpage each week.

Related Courses

Here is a list of similar/related courses offered by other instructors, sorted into several categories based own their flavors.

- *Metric Embeddings* by Avner Magen
- *Metric Embeddings and Algorithmic Applications* by Moses Charikar
- *Succinct Graph Structures and Their Applications* by Merav Parter
- *Metric Embeddings* by Uli Wagner and Jiří Matoušek
- *Algorithmic Applications of Metric Embeddings* by Anupam Gupta and R. Ravi

Teaching Methods and Expectations

The class will mostly be ***lecture-based*** in a seminar style. The instructor and the students will take turns to present and to scribe. There will also be ***reading assignment*** and ***homework problems*** to complement the lectures. Every student is expected to finish the reading assignment before the class, taking notes and contributing in discussion during the class, and solidify the understanding of the material by solving the homework problems.

We will assign roles to each student (technician, historian, architect, etc.) so that you may pay attention to part of the material during the paper reading and provide help and comments during the lecture. And we will use the scribe notes and homework submissions to determine if the students are actively engaging in learning.

Collaboration Rules

You are *strongly encouraged* to use all the resources online/offline. You are also encouraged to work with other groups (and also people outside the course) and discuss your progress, as long as in the end each group individually writes down the solutions in their own words, and properly cites everyone and everything you have consulted. In particular, if you find a solution manual online that contains an answer to the problem, cite the solution manual. If chatting with your friends at gaming night gives you an idea to solve the problem, cite your friends. If your cat walks across your keyboard and sends you to a Wikipedia page that contains a crucial construction to the proof, cite your cat. However, you don't have to cite the lectures and any course related materials. Please refer to the Academic Honor Principle and Sources and Citations at Dartmouth if you have doubts, or contact The Office of Community Standards & Accountability.

Grading

The final grade will be based on presentations, class participation, homework grades, and potentially an optional *research project*.

The tentative weights are 50% on presentation and scribe work, and 50% on homework. If the student choose to do project it will serve as substitute for the homework. Because we are likely to have a small-size class, I would prefer to talk to each of you and set up concrete goals after reviewing your submission to homework 0. The grading will also take into account the difference between student groups (undergrads vs graduates).

Tentative Course Schedule and Assignments

All materials are subject to change based on the actual class interaction and student feedback.

Theory of Metric Embeddings

- Week 1: ***Metric Embeddings***
  - metric space, distortion, p-norms, probabilistic embedding
  - isometric embedding between p-norms, lower bounds
- Week 2: ***Dimension Reduction***
  - Johnson-Lindenstrauss lemma, random projection
  - Bourgain's theorem for embedding into Euclidean spaces

Sketching Planar Metric

- Week 3: ***Planar Graphs***
  - Monge property, spanner, emulator, quadratic-size emulator
  - separators, embedding into Euclidean space
- Week 4: ***Tree Metrics***
  - shortest-path separator decomposition, tree-covers
  - embedding into small-treewidth graphs with additive-distorion
- Week 5: ***Algorithmic Applications***
  - One-hole metric, compression via VC-dimension, approximate diameter
  - Portals, sub-linear-size planar ε-emulator
- Week 6: ***Embedding into ℓ_1***
  - Sparsest cut, multi-commodity flow, flow-cut duality; Okamura-Seymour isometric embedding
  - padded decomposition, $O(\log n)$-approximation for the sparsest cut, planar embedding conjecture

Other Compression Schemes

- Week 7: ***Labeling Scheme***
  - Separator-based labeling scheme, $\sqrt{n}$ upper bound and $n^{1/3}$ lower bound on size
- Week 8: ***Distance Oracle***
  - Exact distance oracles
  - Approximate distance oracles
- Week 9: ***Topics***
  - ARV; inapproximability in ℓ_1, streaming and sketching, compressed sensing, expanders

Religious Observances

Some students may wish to take part in religious observances that occur during this academic term. If you have a religious observance that conflicts with your participation in the course, please meet with me before the end of the second week of the term to discuss appropriate accommodations.

Student Accessibility and Accommodations

Students requesting disability-related accommodations and services for this course are required to register with Student Accessibility Services (SAS; Getting Started with SAS webpage; student.accessibility.services@dartmouth.edu; 1-603-646-9900) and to request that an accommodation email be sent to me in advance of the need for an accommodation. Then, students should schedule a follow-up meeting with me to determine relevant details such as what role SAS or its Testing Center may play in accommodation implementation. This process works best for everyone when completed as early in the quarter as possible. If students have questions about whether they are eligible for accommodations or have concerns about the implementation of their accommodations, they should contact the SAS office. All inquiries and discussions will remain confidential.