CS 784: FOUNDATIONS OF DATA MANAGEMENT

Fall 2019
INTRODUCTION

• undergrad in Athens, Greece
• Ph.D. in University of Washington (the other UW)
• at UW-Madison since 2015!

Research Interests

• parallel query processing
• data pricing
• uncertainty in data management
Course Logistics
COURSE FORMAT

• Lectures **Tu+Th** 2:30-3:45 pm @ CS 1263

• Office Hours: **Tu** 1:30-2:30pm or by appointment


• Mailing List: compsci784-1-f19@lists.wisc.edu
COURSE STRUCTURE

The course will have two parts:
1. Query Languages + Complexity
2. Advanced Topics: provenance, privacy, uncertainty, stream processing, ...

The lectures will be on the blackboard. For some lectures I will post notes on the webpage, for others we will focus on specific papers
PREREQUISITES

It will be helpful if you have good knowledge of:

• Databases, SQL, Relational Algebra
• Algorithms
• Complexity
GRADING

• Class participation: 15%
• Homework (3): 30%
• Paper reviews (4): 15%
• Research project: 40%
HOMEWORK

• Individual assignments
• Submitted through Canvas (use Latex!)
• You can use up to 5 late days for all 3 assignments
PAPER REVIEWS

• Read an assigned paper before the lecture
• Submit a brief review of the paper
• Answer a few questions related to the content of the paper
RESEARCH PROJECT

• In groups of 1-3
• Independent research on any topic related to the course
• Deliverables:
  – 9/29: email groups + tentative ideas
  – 10/13: project proposal
  – 11/10: milestone
  – 12/5+12/10: project presentations (10% of grade)
  – 12/15: final report
SAMPLE PROJECTS

• A Lightweight Approach to Approximately Query Big Data
• Efficient Multiway Joins on Heterogeneous Parallel Networks
• Materialized Views In Data Warehousing Environments
• Implementing Datalog on an Asynchronous Distributed Dataflow Framework
What is This Class About?
WHAT IS THIS CLASS ABOUT?

- Data is everywhere!
- Managing data is critical:
  - scientific discoveries
  - online services (social networks, online retailers)
  - decision making
- **Databases** are the core technology
- **In this class:**
  - Foundations of data management
CLASSIC DATABASE THEORY

- Conjunctive Queries
- Query containment/equivalence
- Query complexity
  - how fast can we evaluate a join?
  - how big can the result of a join be?
  - are some join queries easier to compute than others?
Datalog is a declarative language that allows us to express larger classes of queries!
QUERY EVALUATION

• How do we evaluate queries in parallel environments?
  – MapReduce
  – Spark
• How do we evaluate queries in streaming environments?
How do we deal with uncertain data?

- probabilistic databases
- consistent query answering
- repairs
OTHER TOPICS

- Provenance
- Differential Privacy