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 1221

• Office Hours: **Tu** 4:00-5:00pm or by appointment

• Webpage: [http://pages.cs.wisc.edu/~paris/cs784-s19/](http://pages.cs.wisc.edu/~paris/cs784-s19/)

• Mailing List: compsci784-1-s19@lists.wisc.edu
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

Not any formal prerequisite. 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 (5): 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:
  – 2/9: email groups
  – 2/25: project proposal
  – 4/1: milestone
  – 4/30+5/2: project presentations (10% of grade)
  – 5/5: 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!

LogicBlox

bdddbdddb

OVER ONE QUATTUORDECILLION RELATIONS SERVED
QUERY EVALUATION

• How do we evaluate queries in parallel environments?
  – MapReduce
  – Spark

• How do we evaluate queries in streaming environments?
UNCERTAIN DATA

How do we deal with uncertain data?

• probabilistic databases
• consistent query answering
• repairs
OTHER TOPICS

- Stream Processing
- Provenance
- Differential Privacy