CS 3950
Introduction to Computer Science Research

MEETING 2: SCIENCE, “COMPUTER SCIENCE,” AND RESEARCH

Outline

- What is science?
- What is Computer Science?
- What is CS research?

Outline

- **What is science?**
  - Scientific Method
- What is Computer Science?
- What is CS research?

What is science?

What is science?

A process that systematically builds and advances knowledge by explaining and predicting phenomena in the real world

Science according to Bacon

Something happens in the world

Form a hypothesis that explains/predicts it

Test the hypothesis through experimentation

If successful, use it as a model

- Until counterexample disproves it

Scientific Method in more detail

1. Start with a question or observation
2. Learn about it
3. Propose explanation for natural phenomenon
4. Proposal is a prediction, hypothesis
   - When using if/then, if is the theory/mechanism being tested,
     then is the empirical findings
5. Design a way to test hypothesis
6. Experiment to gather data
7. Look at data, what does it mean, can we say they support hypothesis?

Scientific Method

The data supports the hypothesis!

ARE WE DONE?

NOPE.

Scientific Method

New hypotheses, new experiments to retest ideas

- Multiple conditions, times, people -> well-supported hypothesis -> model -> theory

Meetings, publish, share

- Value of skepticism, criticism, peer review

Become part of the scientific literature

The Scientific Method is a way of thinking, not a formula

Outline

- What is science?
- **What is Computer Science?**
  - Main branches
  - Principles
- What is CS research?

What is Computer Science?

Denning:

> Science of information processes and their interactions with the world

What is Computer Science?

Some combination of:

- Science
- Engineering
- Mathematics

But wait, isn't all "Computer Science" science?

"Computer Science" is a name of a field

Computer science is the scientific endeavor relating to computing

Computer science

Examples using hypothesis testing

Engineering or “computing arts”

Useful practices of a field, often **application** of science and math

Mathematics

Studying numerical, space, time properties of information processes

- Can be abstract or applied

Outline

- What is science?
- **What is Computer Science?**
  - Main branches
  - **Principles**
- What is CS research?

Principles

Computation (meaning and limits of computation)

Communication (reliable data transmission)

Coordination (cooperation among networked entities)

Recollection (storage and retrieval of information)

Automation (meaning and limits of automation)

Evaluation (performance prediction and capacity planning)

Design (building reliable software systems)

http://denninginstitute.com/pjd/GP/gp_overview.html

Example problems:

Computation
- What is computable? How “tractable” is a computation problem?

Communication
- How should we encode and transmit data, efficiently and reliably?

Coordination
- How do we ensure multiple interacting computing systems agree on a computation?

Recollection
- How do we organize information for efficient storage and retrieval?

Example problems

Automation

- To what extent can we automate/replicate behaviors of systems in nature?

Evaluation

- What is the expected behavior of a computer system under various inputs/conditions?

Design

- What programming models and abstractions lead to efficient, reliable programs?

Computer science in science

Bioinformatics

Experimental particle physics

Quantum computing

Political and social science

Protein folding

Public health informatics

…

Summary

There is science in Computer Science
- Not all Computer Science is science

CS combines science, engineering, and math
- Hypothesis formulation and testing
- Design and implementation of computer systems
- Proofs and analysis of algorithms and information transformation

CS has been combined with nearly every other field
- Nearly endless combinations leading to nearly endless interesting questions to explore

Outline

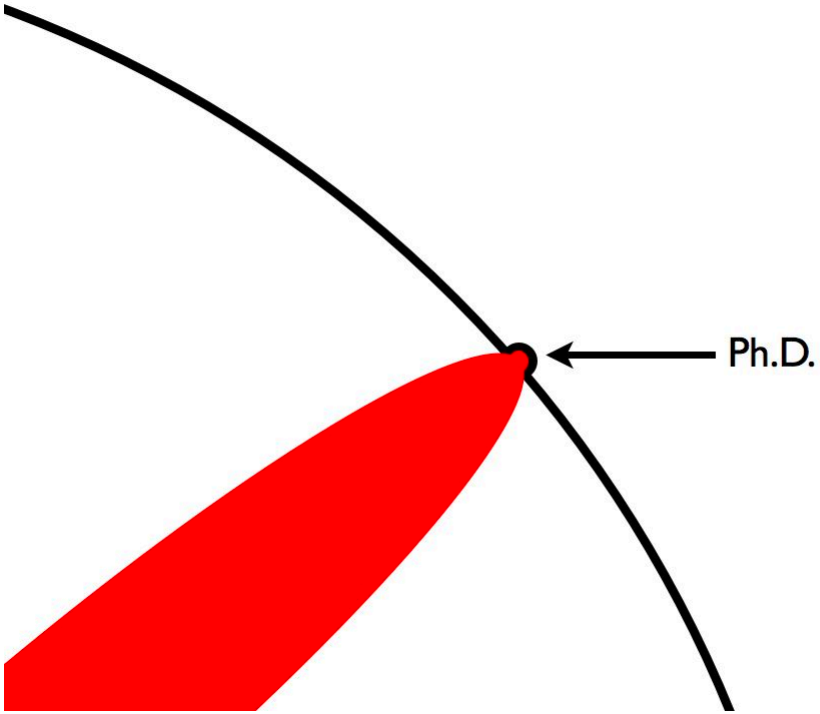
- What is science?
- What is Computer Science?
- **What is CS research?**

What is CS research?

Pushing the boundary of knowledge

http://matt.might.net/articles/phd-school-in-pictures/

What is CS research?



CS Research Areas

ALGORITHMS & THEORY

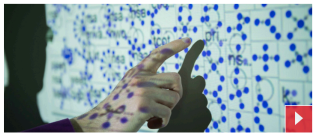
algorithmic game theory, approximation algorithms, complexity theory, convex optimization, cryptography, data privacy, distributed computing, learning algorithms, pseudo-randomness

VIEW GROUPS

ARTIFICIAL INTELLIGENCE

user modeling, intelligent tutoring systems, mixed initiative design, computational creativity, computational psychology, computer vision, affective computing, human behavior modeling

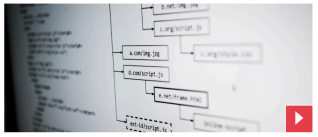
VIEW GROUPS



DATA SCIENCE

machine learning, statistics, data mining, parallel & distributed data analysis, database systems, information retrieval, knowledge representation, information visualization, natural language processing, computational biology and bioinformatics, computational social science, digital humanities, health informatics, business and predictive analytics

VIEW GROUPS



FORMAL METHODS

automated theorem proving, concurrency, decision procedures, formal verification, model checking, numerical procedures, refinement, reliable systems

GAMES

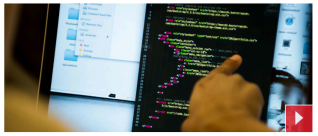
procedural content generation, behavior or user modeling in games, computational narrative, game human interaction, game analytics, game user research, context-aware games, games for impact, games as crowdsourcing, designer support AI tools, behavior modeling of human and team behavior from game data

VIEW GROUPS

HUMAN-CENTERED COMPUTING

conversational agents, multimodal interaction, crowdsourcing, user experience assessment, ubiquitous computing, visualization, computer-supported cooperative work (CSCW)/social computing

VIEW GROUPS



MACHINE LEARNING

graphical models, learning-to-rank, deep learning, semi-supervised learning, health informatics, machine learning theory

VIEW GROUPS

NATURAL LANGUAGE PROCESSING AND INFORMATION RETRIEVAL

information retrieval, information retrieval models, information extraction, sentiment analysis, social media analysis, political text analysis, crowd-sourcing, conversational agents, syntactic and semantic parsing language, machine translation

VIEW GROUPS

NETWORK SCIENCE

complex networks, information networks, political networks, social networks, social network analysis and mining, graph mining, network classification, collaborative decision-making

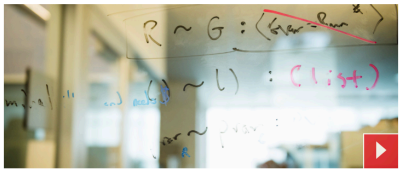
VIEW GROUPS

CS Research Areas

PERSONAL HEALTH INFORMATICS

tele-monitoring and home interventions, mobile health, behavioral modeling, health games, bio-sensing, conversational agents, health education, health behavior change, consumer informatics, assistive technology, health data visualization

VIEW GROUPS



PROGRAMMING LANGUAGES

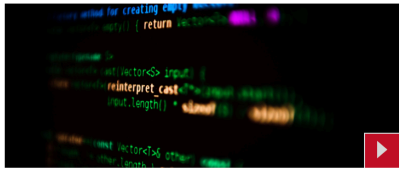
verification, static analysis, compilers, interpreters, memory management, virtual machines, type systems, semantics, domain-specific languages

VIEW GROUPS

ROBOTICS

robotic manipulation, perception, humanoid robotics, human-robot interaction, health counseling robots

VIEW GROUPS



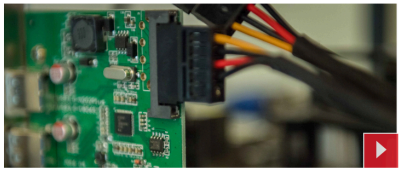
SECURITY AND PRIVACY

cryptography, privacy, system security, software security, mobile and wireless security, network security, distributed systems security, verifiable attack countermeasures, cyber-physical systems security, security analytics

VIEW GROUPS

SOFTWARE ENGINEERING

testing, verification, methods for software development, model-based development, software architecture



SYSTEMS AND NETWORKING

cloud computing, cyber-physical systems, distributed systems, mobile and wireless systems, networking, parallel computing, security and privacy, storage systems

VIEW GROUPS

Northeastern University

DATA Lab @ Northeastern

Scalable Management and Analysis of Big Data

Home | People | Research Opportunities | Recent Publications | Activities

DATA LAB @ NORTHEASTERN

The Data Lab @ Northeastern University is a team of faculty and students who explore a range of research problems in scalable data management and analysis. Our work ranges from fundamental questions on the complexity of data management problems to practical applications with domain scientists and covers areas such as large-scale and parallel data analysis algorithms, graph data management, and uncertain data. We participate in a number of interdisciplinary research projects and collaborate with other faculty at Northeastern and database groups across the world. And we are growing!

What is CS research?

Projects led by one or more PIs
- Typically faculty, with Ph.D.

May span more than one area

Project team may include
- Ph.D. students
- Postdoctoral associates
- Undergraduate students
- Developers

https://www.khoury.northeastern.edu/research/research-projects/

How does CS research get done?

Can be done (mostly) solo, but is often a team effort

May involve

- proving theorems
- writing code
- designing a programming language
- building physical systems
- interacting with people

Often requires some kind of evaluation and/or analysis

Needs to be effectively communicated and subject to critical review

- We typically publish in peer-reviewed conferences and journals
- Lead author usually gives oral presentation

What are products of CS research?

Peer-reviewed, published paper

- Gold standard, look for well-known conferences/journals
- “peer reviewed” doesn’t always mean “correct”
- arXiv, not peer reviewed, often used for flag-planting

Artifacts

- Systems, e.g., software (hopefully open source) and hardware (e.g., robots, chips, …)
- Data and analysis code
- Designs, visualizations, …

Theories/Models/Accepted Knowledge

Where do you find CS research

Google (Scholar)

DBLP (https://dblp.uni-trier.de)

Khoury website (khoury.northeastern.edu)
- Search by area/project
- Visit faculty pages

ACM, USENIX, and IEEE conferences/journals
- Access from NEU network to get around paywalls
- Which ones to focus on? CS Rankings can help (https://csrankings.org)

Arxiv
- Wild west of papers, some areas highly respected, others not
- Often a good way to find papers outside a paywall

Homework 1

Computer Science research “field trip”

- Go to https://www.khoury.northeastern.edu/res
- Pick a research area that interests you
- Scroll to the bottom for PhD students
- Visit a few students to understand what they wo
- Describe the field in terms
  - Is it science, engineering, math, s
  - What are example of problems w
- Do this for a second resea

Anurag Bhardwaj
Part-Time Lecturer - Silicon Valley

Rukmini Vijaykumar
Part-Time Lecturer

Albert Cheu
PhD Student

Matthew Dippel
PhD Student

BIOGRAPHY

Lucianna Kiffer is a PhD student in the Algorithms and Networks program at Northeastern University’s College of Computer and Information Science, advised by Professors Alan Mislove and Rajmohan Rajaraman. A native of Belo Horizonte, Brazil Lucianna earned her bachelor’s degree in both mathematics and computer science from Tulane University.

Lucianna’s research areas include Algorithms and Theory and Systems and Networks. She is a member of the Social Networks Group. Lucianna has an interest in peer-to-peer networks and is currently working on a project that utilizes blockchain technology.

EDUCATION

- BS in Mathematics and Computer Science, Tulane University

About Me

- Field of Study: Algorithms and Networks
- PhD Advisors: Alan Mislove and Rajmohan Rajaraman

What are the specifics of your graduate education (thus far)?
Professor Rajaraman, Professor Mislove and I have begun a project exploring Ethereum; a blockchain-based decentralized virtual machine. At this point in the research we are generally looking into scalability concerns involving Ethereum’s network, measurement based questions about the network, and generally what can be built on the Ethereum network and networks like it.

What are your research interests?
I am interested in peer-to-peer networks and distributed computing both in application and theory. Computer science was a single class I had intended to take my freshman year of college while I studied chemistry and it somehow became what I want to study for the rest of my life, while math has always been a passion of mine.

What’s one problem you’d like to solve with your research/work?
I would like to work on a solution to one of the scalability concerns of blockchain networks/technologies.

What aspect of what you do is most interesting?
One of the most interesting aspects of distributed networks for me is the amount of information of the overall network that can be propagated through minimal communication and/or in scenarios where no trust is assumed between parties in the network.

Homework 1

Will be posted on website this evening

- Read instructions carefully

Due at beginning of class Thursday

- Bring a copy, you will need it
- Submit electronically