# Database Metatheory: Asking the Big Queries

Christos H. Papadimitriou

Slides based on slides from Jane Zhang and Rachel Pottinger

Discussion originally by Mike Wood

# Introduction

- What is theory?  In general:
  - Abstraction: suppression of low-level details
  - Goal: see fundamental truths obscured by details
- In CS, theory is generally mathematical:
  - Developing models
  - Using model
  - Analyzing models

# How does one do theory?

- Develop Mathematical Models
  - E.g., Turing machines
- Propose Complexity-Reducing Solutions
  - E.g., algs. for answering queries using views
- Analyze
  - E.g., transaction processing
- Explore!
  - What are real semantics of NULL?

# Discussion (1 - part 1)

- "…nowhere is this adaptation to the environment more prevalent and complexity-inducing than in databases, whose purpose is to represent parts of the environment, as well as to interact with other parts."
- With your neighbor, discuss…
- What does the author mean by representing and interacting with the environment? Which aspects of database do you see as being representative and which as being interactive?

# Discussion (1 - part 2)

- Again, with your neighbor, discuss…
- Does the representativeness or interactiveness of an aspect of databases change depending on the underlying data model?
  - relational vs. object oriented vs. XML
- Consider both internal and external aspects
  - Internal: query processing, transactions, etc.
  - External: query language, result set, etc.

# The Joys and Pains of Exploration

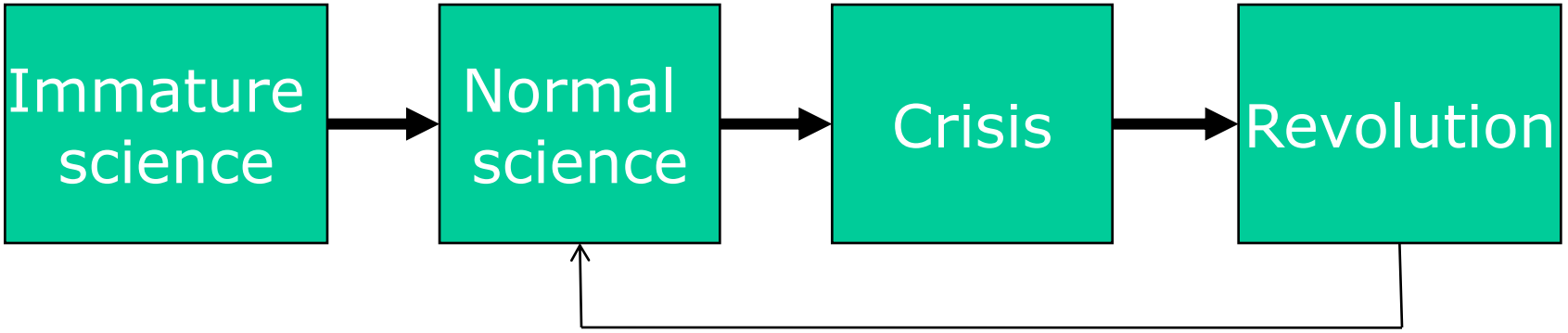
- Joys:
  - Historically useful
  - In reasonable amounts, ensures good health
  - Theories are pretty: people will do it anyway
- Pains:
  - Must not consistently ignore practice
  - Requires careful exposition of relevance and applicability
  - Too much can lead to crises

# What is “Good Theory”

- *All* ideas improve knowledge
- But whether it’s “good” theory largely depends on propaganda
  - Needs to influence beyond itself
  - Has to at least *be able to* influence practice

The ultimate influence: launching a victorious scientific revolution

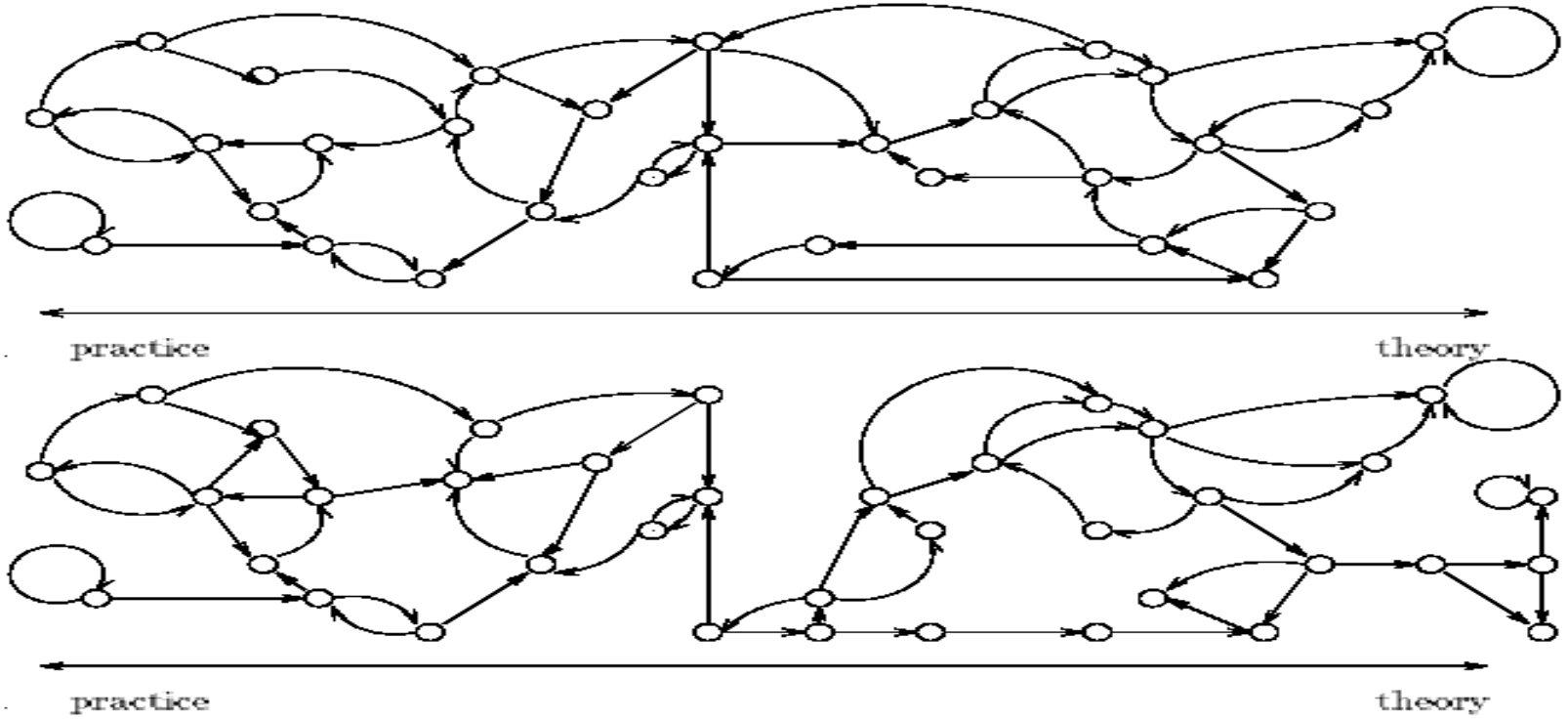
# On Paradigms and Revolution (Thomas Kuhn's Model)




- "Normal" science has a predominant *paradigm*
  - Scientists pressured to defend paradigm and show it works
- Eventually, a crisis causes a revolution
  - E.g., relational model

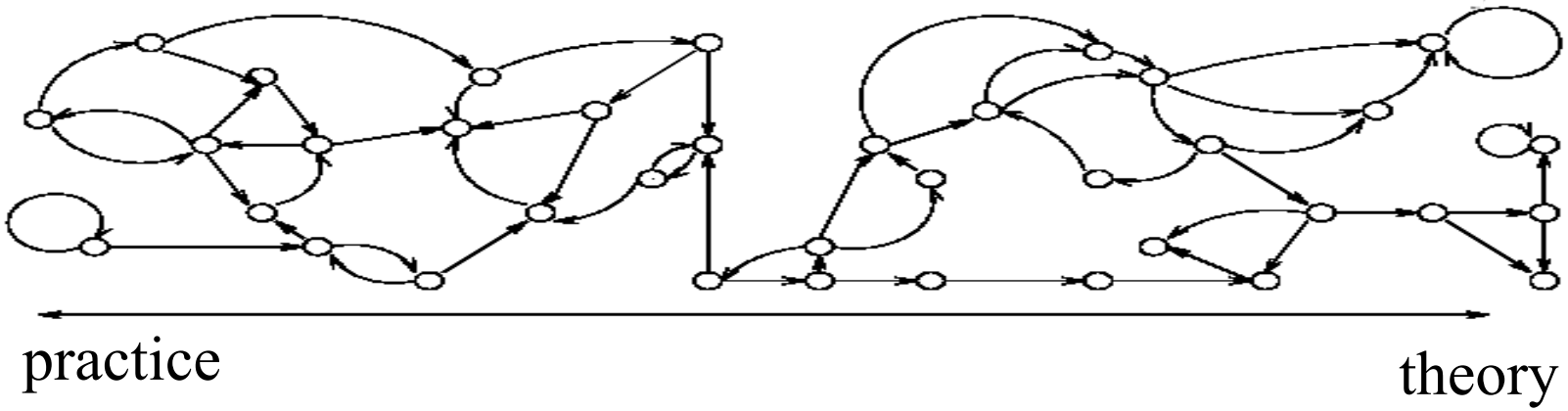
What's theory's role?

# Theory's role in revolution: normal




- Lots of connections
- Most theory within a few hops of practice, and vice-versa

# Theory's role in revolution: crisis




- Long paths from theory to practice
- Some nodes have no or little routes to practice
- In short term, this is very bad
- In *long* term, can help create new paradigm and new practice

# What about database theory? (as seen by PODS papers)

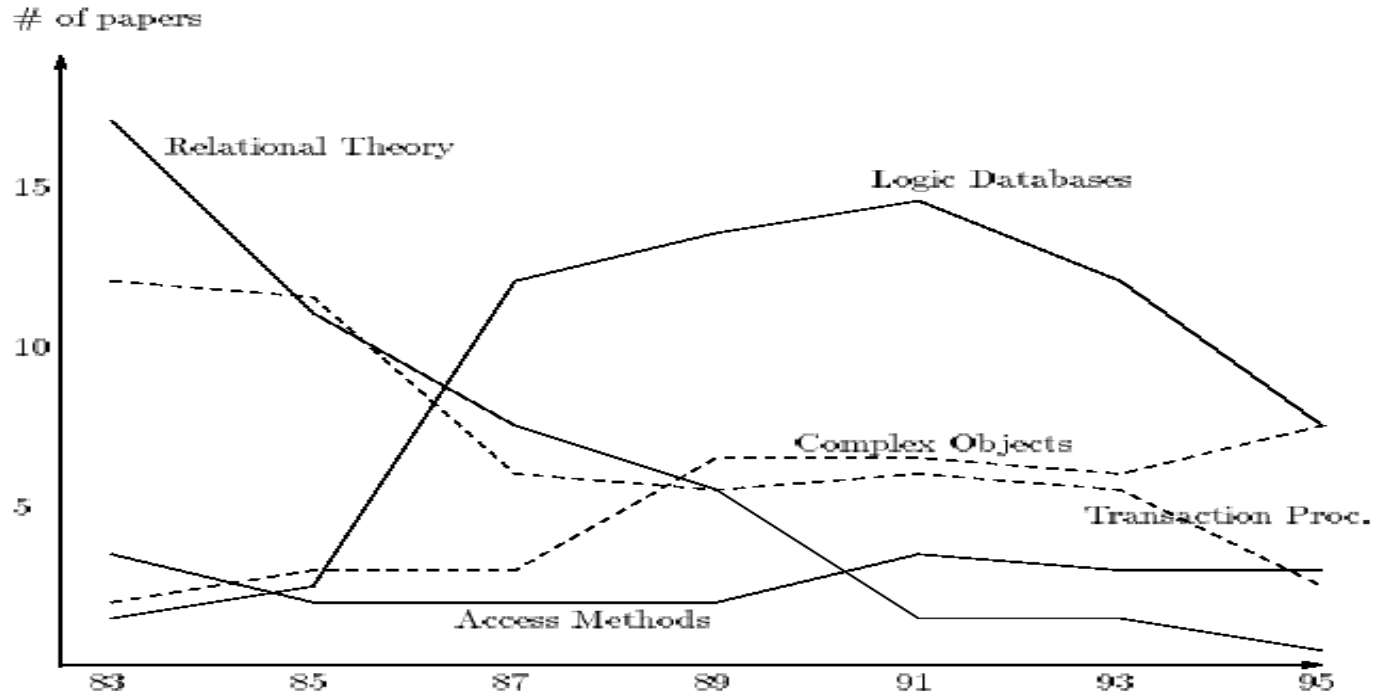



Figure 3: The number of PODS papers in five areas, averages for the two-year period ending in the year indicated.

- In the beginning (1982), there was relational theory and transaction processing
- Then datalog, objects, XML (not shown)

# Discussion (2 - part 1)

- With respect to Kuhn's model, in what state is each of the following areas of database research? Normal? Crisis? Revolution? Justify your choice.

  - relational roots
  - query optimization
  - query execution
  - transaction processing
  - extensible databases
  - distributed databases
  - views
  - adaptive execution
  - object oriented DBs
  - XML
  - temporal and RT DBs
  - data mining
  - streaming data
  - DB administration

# Discussion (2 - part 2)

- With respect to Kuhn's model, what state is database research as a whole currently in? Normal? Crisis? Revolution?

# How did database theory do?

- Big Win:
  - Relational model & normal forms
- Big Loss:
  - Datalog & recursive queries (a bit better now)
- Draws:
  - Object-oriented models?
  - Only simplest concurrency control used

# Christos's Theory Soapbox

- Good: Only now can one become a famous pure theoretician
- Bad: CS Theory is roundly bashed in some areas

And then there's applicability…

# Dangerous Applicability Claims

- Recursive applicability
  - The last n papers said it was applicable
- Remote applicability
  - People in other fields find it applicable
- Applicability by association
  - If X is relevant to Y, then anything involving X must be applicable

# Discussion (3 - part 1)

- Applicability fosters negative cycle, distancing theory and practice communities

- What makes good theory? Scientific merit? Applicability? Propaganda?

# Theory in time of Crisis

- "De-intellectualization" is the order of the day: Research & Academia are logical and strategic targets
- Pride on how pervasive we are → A cacophonous and off-tempo chorus
- Theoretical CS is coming of age:

➢ Basic models have been explored

➢ New models have not had the attention

# What should Theoreticians do?

- Must pay limited attention to the voices of the crisis
- Should not feel obliged to coordinate our research goals with current applied research
- Should question and challenge the prevailing ideology within theory
- Should be even more independent, bold, imaginative, exploratory, anarchistic

# What should Theoreticians do?

- Should focus on complexity reducing program of CS
- Should focus on the connectivity increasing functions of theory

*It is darkest before the dawn*

# Discussion (3 - part 2)

- Is the research community insecure? Should it be? Is industry wrong to demand immediate applicability from research?
- [MSc] Do you feel compelled to conjure up phony applicability and motivation for your projects? Or do you just want to publish something?
- [MSS] What (if anything) do you value from purely theoretical research or research with no immediately clear application?

# Thank You