DATABASE METATHEORY: ASKING THE BIG QUERIES

Christos H. Papadimitriou

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?

THE JOYS AND PAINS OF EXPLORATION

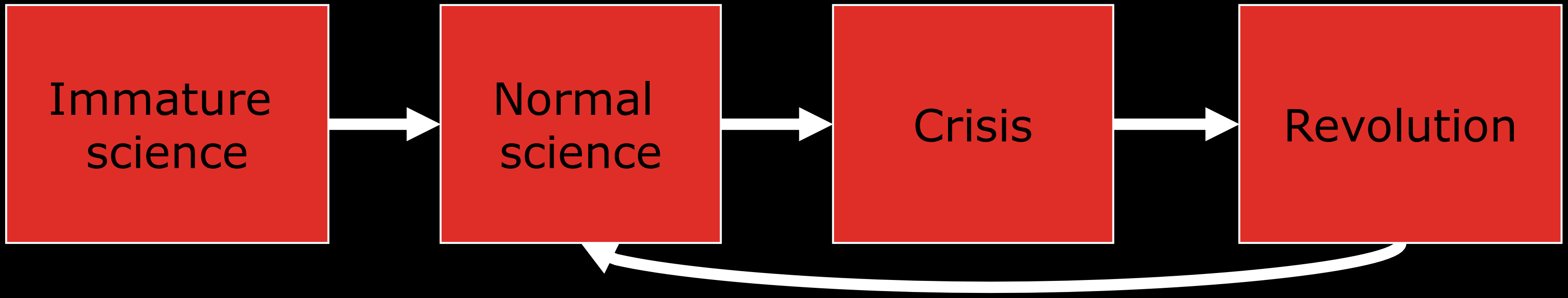
- Joys:
  - Historically useful
  - In reasonable amounts, ensures good field health
  - Theories are pretty: people will do it anyway
- Pains:
  - Must not consistently ignore practice
  - Requires careful exposition of relevance and applicability
  - Too much can lead to crises

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)



Immature science

Normal science

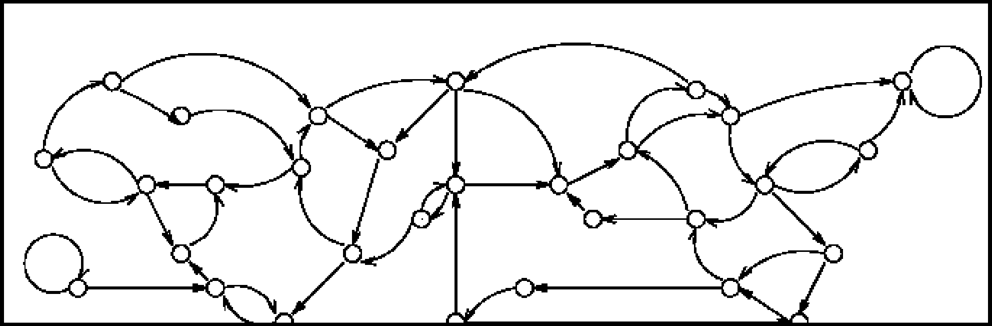
Crisis

Revolution

- "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

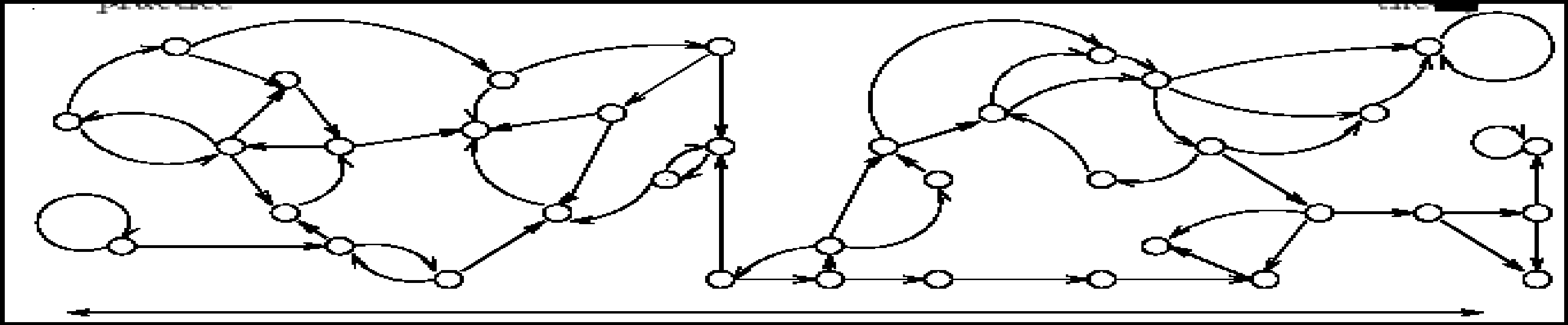


practice

theory

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

THEORY'S ROLE IN REVOLUTION: CRISIS



practice

theory

- 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)

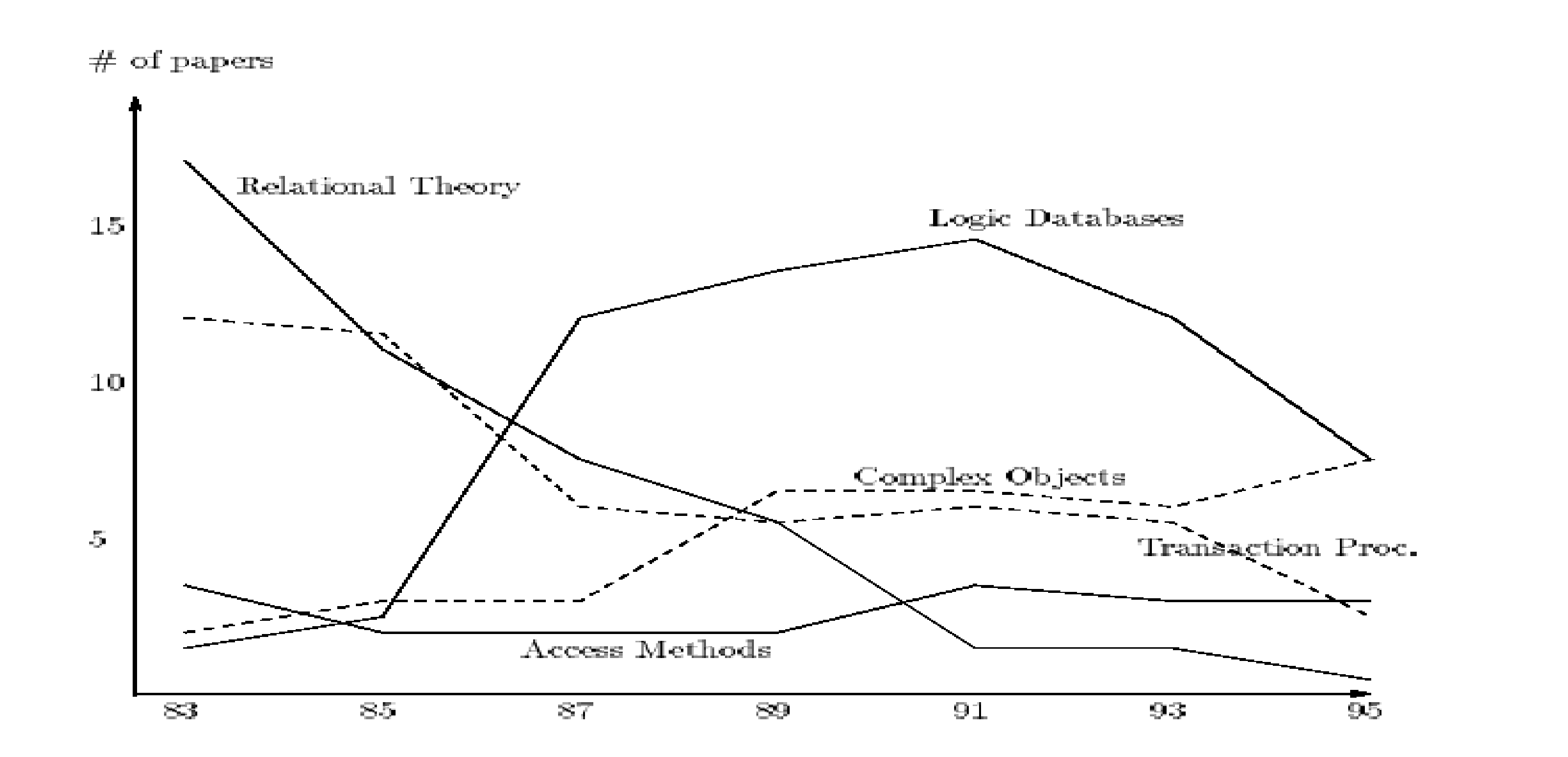


Figure 3: The number of PODS papers in five areas, averages for the two-year period ending in the year indicated.

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

HOW DID DATABASE THEORY DO?

- Big Win:
  - Relational model & normal forms
  - Database design tools
- 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