



DATABASE METATHEORY: ASKING THE BIG QUERIES

Christos H. Papadimitriou

Presenter: Wilson Tu
Discussion lead: Ishita Haque
Date: March 20, 2024

INTRODUCTION

- General public's perception of theoretical CS
- What is theory broadly?
 - Significant abstraction to scientific research
 - Suppression of low-level details
- What is theory more specifically?
 - Sophisticated math
 - Developing models
 - Using model
 - Analyzing models
 - Cure complexity of models

QUESTIONS TO KEEP IN MIND

- Relationship between theory and practice?
- How can theory help us?
- Comparison of the CS field with other sciences in terms of theory?

HOW DOES ONE DO THEORY?

- Develop Mathematical Models
 - E.g., Turing machines, Relation model
- Propose Complexity-Reducing Solutions
 - E.g., Algs. for answering queries using views
- Analyze the mathematical model
 - E.g., Importance of normalization in relational DB
- Explore!
 - “Anything goes!”

THE FORs AND AGAINSTs OF EXPLORATION

- Fors:
 - Historically useful
 - In reasonable amounts, ensures field's health
 - Theories are pretty: people will do it anyway
- Againsts:
 - Disorient the field
 - Will not thrive if ignoring practice
 - Requires honesty and discipline

HOW ARE THEORY WORKS JUDGED?

- Criterion for theorem in CS
 - Elegance
 - Depth
 - Advances in complexity-reducing program
 - Or points out a setback
- Results of all theories
 - Self-contained negative results
 - Potential to be positive result (invitation for experiments)
 - “Negativity” in the eye of the beholder

Discussion (in pairs)

'“Negativity” is in the eye of the beholder’.

- Can you think of examples in data management, or other research areas, where a negative result led to a significant breakthrough or shift?
- How can researchers, as a community, be more accepting of negative results?
- How can researchers be encouraged to share their negative results?

WHAT IS “GOOD THEORY”

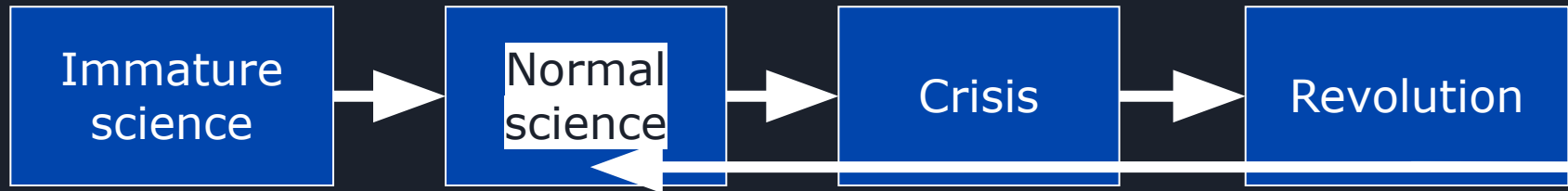
- All ideas improve knowledge. “Anything goes”
- There is no “bad science” but being successful is another category
- “Successful” theory largely depends on propaganda
 - Needs to influence beyond itself
 - Has to at least *be able to* influence practice
- Implications
 - Exploration is legitimate
 - Successful theoretician -> popularizer

The ultimate influence: launching a victorious scientific revolution

Discussion (in groups of 4)

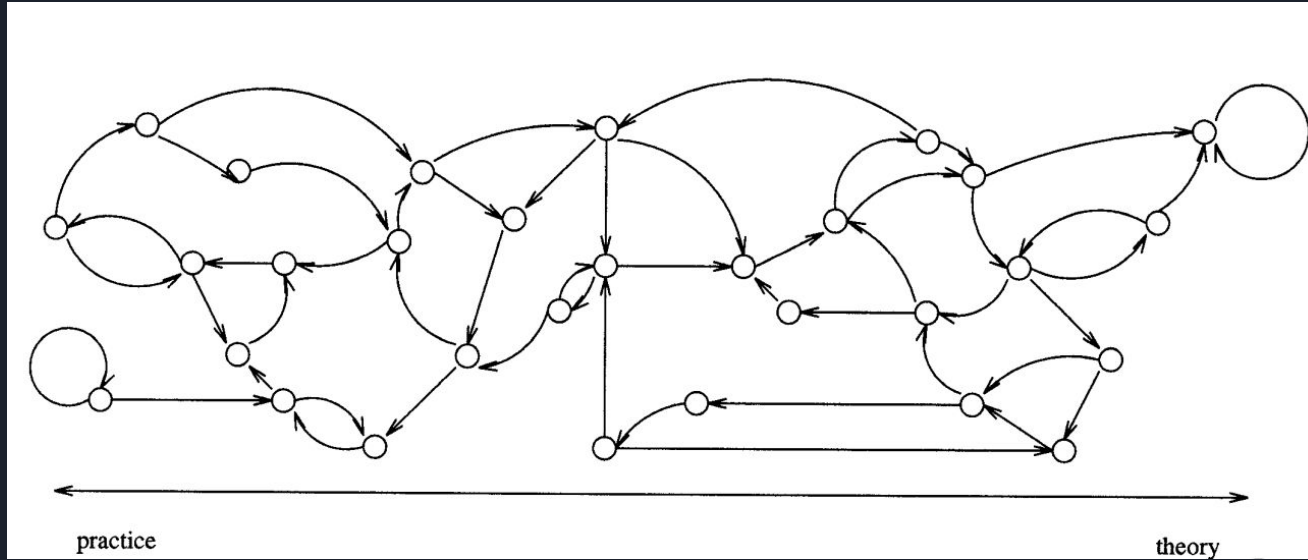
- *"Theoretical research success will depend mainly on its propagandistic value, on its ability to contaminate its environment, especially on its potential to influence practice."*
- Needing to propagandize is common in CS research, and research in general.
 1. Do you agree with the author's interpretation of "good and successful theory"? Reasons?
 2. Do you think that success in theory is based more or less on propaganda than in general? Why or why not?
 3. How can you relate this need for propagandizing to other research we read previously, or to your own research fields?

ON PARADIGMS AND REVOLUTION (THOMAS KUHN'S MODEL)



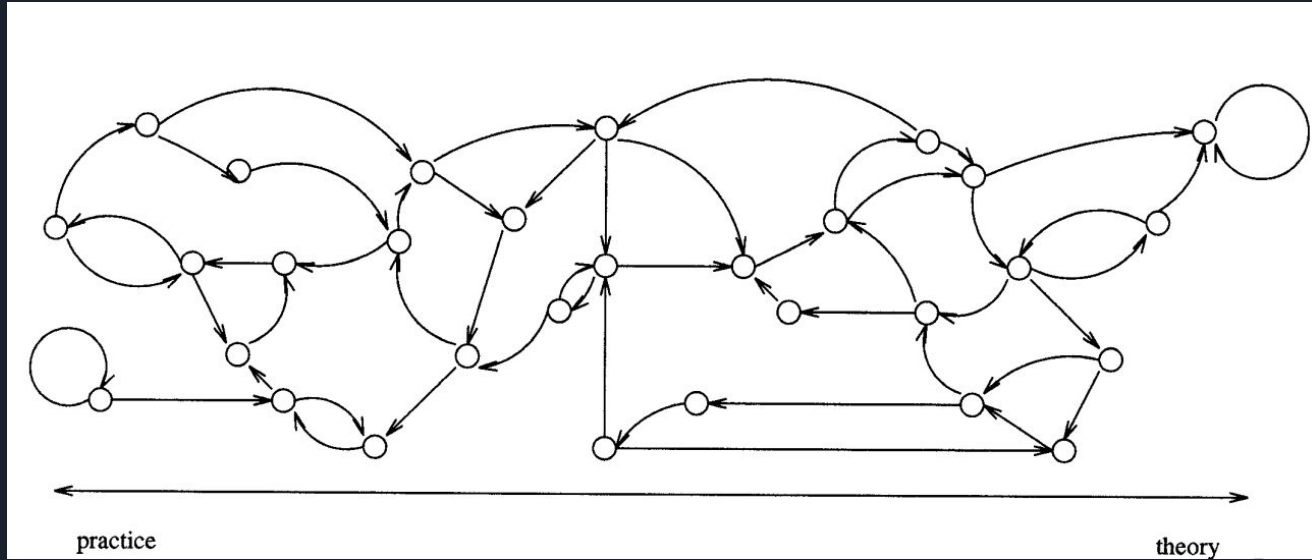
- “Normal” science has a predominant *paradigm*
 - Scientists pressured to defend paradigm and show it works
- Eventually, a crisis causes a revolution
 - E.g., relational model
- This is specific for natural sciences, what about CS, a more “dynamic field”?
- Objective reality in natural sciences, lack of it in CS

THEORY'S ROLE IN REVOLUTION: NORMAL



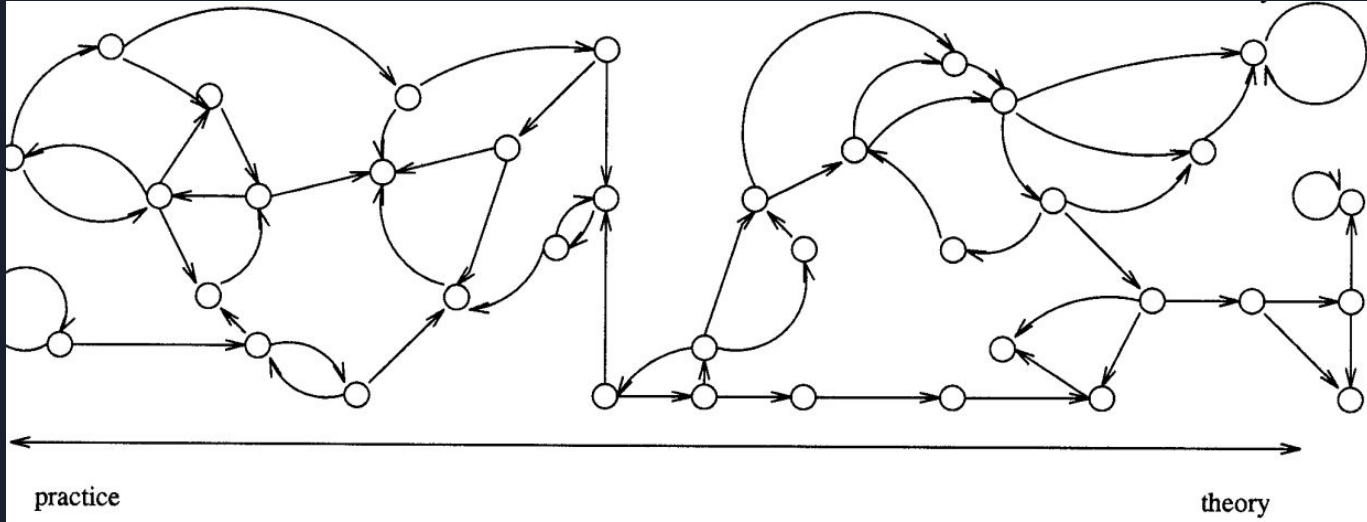
- The interpretation of the nodes are left ambiguous
- The edges show connections between different ideas

THEORY'S ROLE IN REVOLUTION: NORMAL



- Lots of connections from both theory and practice
- Most theory within a few hops of practice, and vice-versa
- Few isolated clusters from both sides

THEORY'S ROLE IN REVOLUTION: CRISIS



- Only subtle global differences
- 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

SOUND FAMILIAR?

- Relational model (Codd's paper)
 - It was a powerful and attractive proposal.
 - It was explicitly open-ended, a whole framework for research problems, applications, and experiments;
 - It came as the result of a crisis (or was it immature science?)
 - It was indeed followed by a period of normal science.
- At the time of the paper it was unclear whether the normal science period has ended or not

Discussion (in groups of 3)

1. The author states that theory research was at a time of crisis in 1995. What stage would you argue data management research is in now? If "normal," what do you think would be required to get data management out of this state?
 - a. You can look into recent database conference's papers (e.g., SIGMOD or VLDB 2023), or DBLP entries and see what their technical details look like
2. Can you think of some other examples following the normal-crisis-revolution flow of Kuhn's model?
 - a. In data management research
 - b. Within your own field

ON DATABASE THEORY (BY PODS PAPERS)

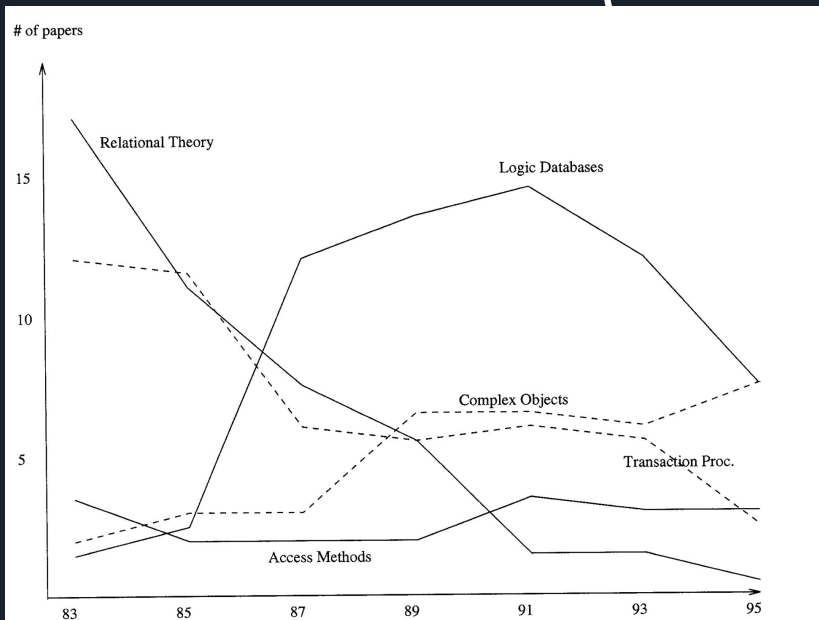


Figure 3: The number of PODS papers in five areas, averages for the two-year period ending in the year indicated.

- In the beginning (1982) relational theory and transaction processing dominated, (little bit of other topics)
- Very dynamic
- Too responsive and subject to fashion to fad

INFLUENCE OF THEORY

- Responds well to applied research
- Major influence:
 - Relational model & normal forms
 - Database design tools
- Things that did not work:
 - Recursive queries
- Minor influence:
 - Contribution to object-oriented models
 - Only simplest concurrency control used

THE HISTORY OF PRACTICE

- Historically theory was prioritized over practice
- Accumulating knowledge was seen as “more superior”
- Practice gained more momentum in the industrial age
- Early CS was dominated by theory, but it became fashionable to criticize theory, and belittle its contribution

DANGEROUS APPLICABILITY CLAIMS

- Fine for but theoretical motivation
- Applicability pitfalls
 - Recursive applicability
 - The last n papers said it was applicable
 - Historical applicability
 - Historically it was important so it's important
 - 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
 - Applicability by pun
 - Using naming to fool audience

Discussion (in groups of 4)

Recursive applicability	The last n papers said it was applicable
Historical applicability	Historically it was important so it's important
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
By pun	Using naming to fool audience

- Which applicability claim would convince you the most? Why?
- Which claim do you think is the hardest to check? How would you assess it?
 - On data management research, or even your own research

CHRISTOS'S THEORY SOAPBOX

- De-intellectualization of academia and research as a whole
- Harassing it's "intellectual vanguards" because field is not matured and lack political entrenchment
- CS was coming of age. New technological challenges do not lead to good theory.
- The solution?
 - Must not listen to the voices of crisis
 - Don't feel obliged to coordinate applied research
 - Be independent, exploratory, and anarchistic.
 - Do your own experiments
 - Don't self pity

Discussion (in pairs)

1. In your opinion, what should be the role of theoreticians in times of crisis? What might be the pros and cons of having an 'independent' approach at these times?
2. In your opinion, what should be the role of applied computer scientists in times of crisis?