HUDK 4050:
CORE METHODS
IN EDM

11/6/18 3:14 PM

In the news

Teachers love Google's education products but are suspicious. Why is a megacorporation giving them a perfect tool for free?

BUSINESS INSIDER

THE JOURNAL

Report: Collection of Ed Data Useful but Challenging

EdSource
HIGHLIGHTING STRATEGIES FOR STUDENT SUCCESS

California needs a new master plan to close the education equity gap

Analytics India MAGAZINE

Finding The Right Mentor In Data Science Is Just As Important As Your Analytics Education

United States of Apathy

2016 U.S. presidential election results if abstention from voting was counted as a vote for "Nobody"

https://urbanobservatory.maps.arcgis.com/apps/MapJournal/index.html?appid=95f38693320c408a87dbb33762d82e82

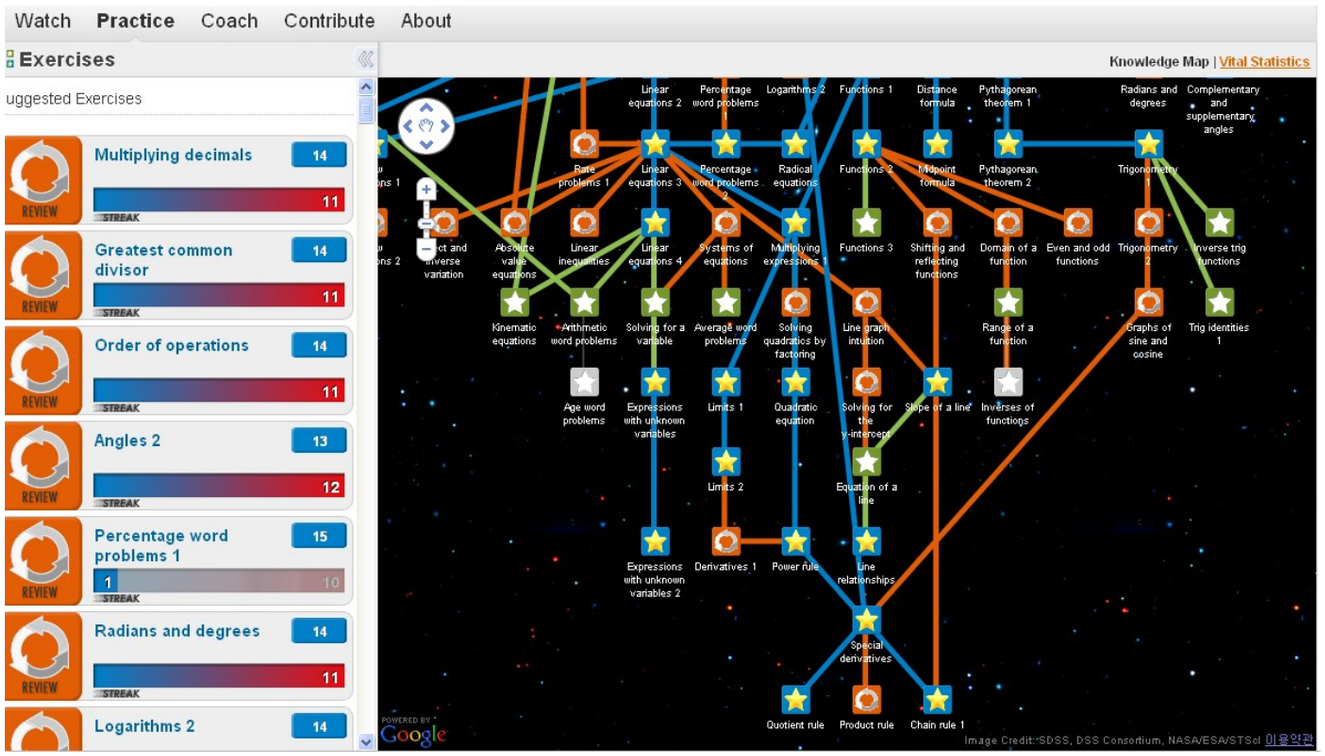
Events

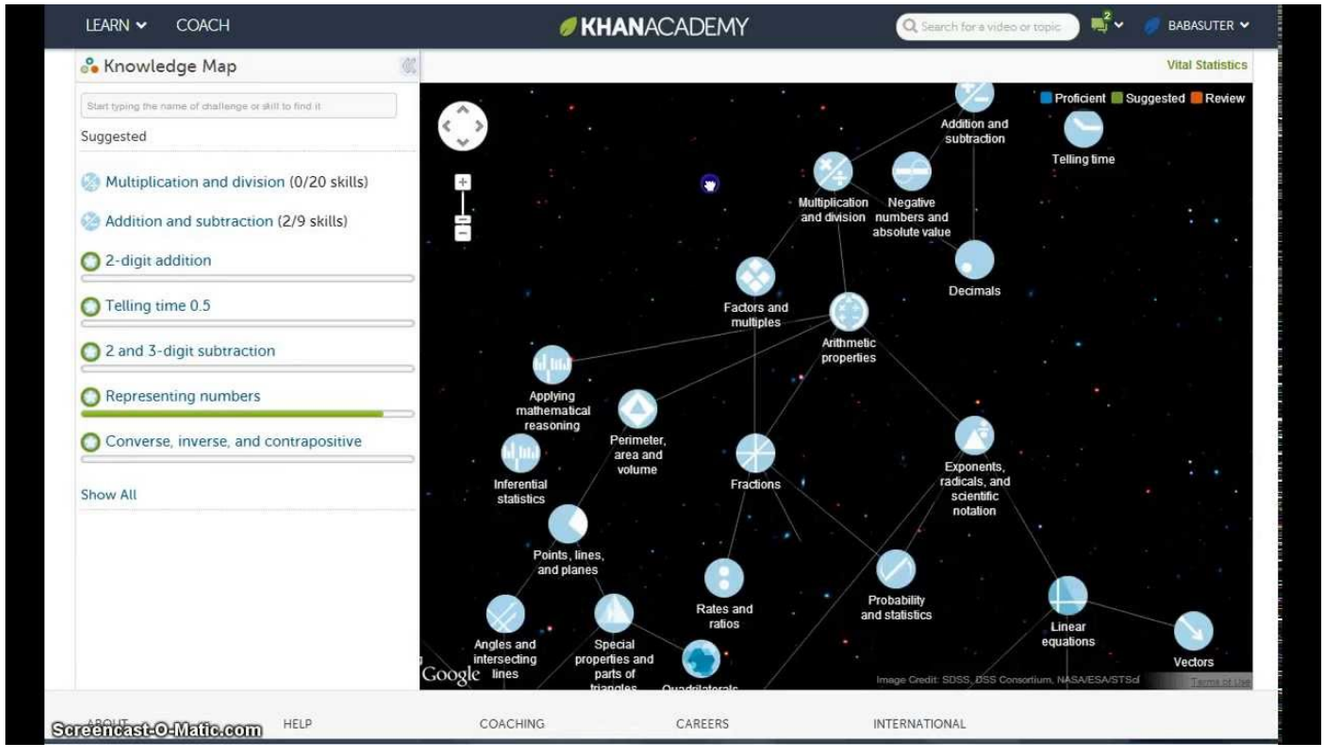
Event	Date	URL
TCLA Social Hour	7:00pm November 7	https://goo.gl/forms/uYVHwVUNT0bnDFbI3
Technical Complexity and Public Discourse: Why It's So Hard to	1:00pm November 9	https://events.columbia.edu/cal/event/eventView.do?b=de&calPath=%2Fpublic%2Fcals%2FMainCal&guid=CAL-00bb9e25-66e323b3-0166-e502e8b1-00002c67events@columbia.edu&recurrenceId=
Data Law in a Global Digital Economy	November 9	https://www.guariniglobal.org/data-law)
NYAS: Deep Learning to Accelerate Drug Development	November 13	https://www.nyas.org/events/2018/deep-learning-to-accelerate-drug-development-an-nyc-symposium/?utm_source=The+New+York+Academy+of+Sciences&utm_campaign=f0807f47cb-eNews_October_2018-10-18&utm_medium=email&utm_term=0_cba25b11d2-f0807f47cb-184577937&mc_cid=f0807f47cb&mc_eid=cfeeec7fb2
People centric approach to optimize Data Science, Commercial impact and Leadership	10:30am November 14	https://events.columbia.edu/cal/event/eventView.do?b=de&calPath=%2Fpublic%2Fcals%2FMainCal&guid=CAL-00bb9e24-655b8449-0165-5e0ea7e9-00001957events@columbia.edu&recurrenceId=
Computing for the Endless Frontier	2:30 November 27	https://events.columbia.edu/cal/event/eventView.do?b=de&calPath=%2Fpublic%2Fcals%2FMainCal&guid=CAL-00bb9e25-66e323b3-0166-e4d9e20e-00002992events@columbia.edu&recurrenceId=
Cross-device User Clustering at Adobe	5:30 November 29	https://events.columbia.edu/cal/event/eventView.do?b=de&calPath=%2Fpublic%2Fcals%2FMainCal&guid=CAL-00bb9e28-655b8cee-0165-5dd5c72b-00001287events@columbia.edu&recurrenceId=
Machine Learning Innovation Summit	December 12-13	https://www.theinnovationenterprise.com/summits/machine-learning-innovation-summit-new-york-2018

Plans

- Today: Assignment 3 (Clustering) Due
- Thursday: Guest Speaker -
- 11/13: Assignment 4 (PCA) Due
- 11/20: Assignment 5 (Prediction) Due
- 11/27: Assignment 6 (CART) Due
- 12/4: Assignment 7 (Diagnostics) Due
- 12/6, 12/11: Assignment 8 work session
- 12/13: Assignment 8 Due and Rating

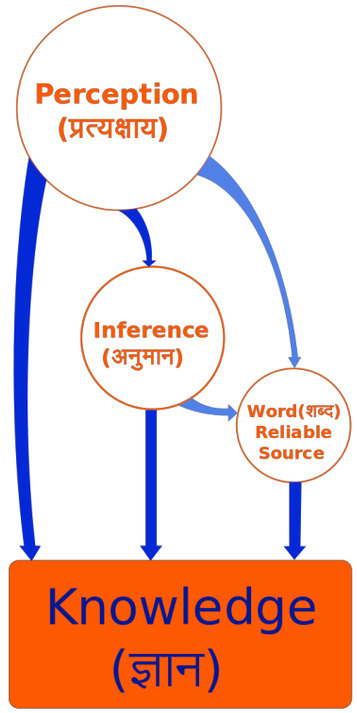
ALICE'S ESTIMATED PROFICIENCY
ON ALL LEARNING OBJECTIVES IN THE GRAPH





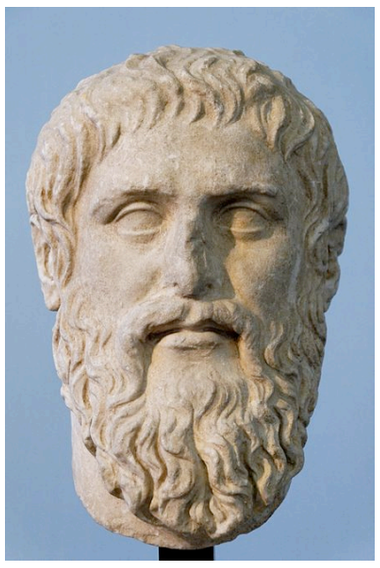
Domain Structure Discovery

- Identifying the structure of knowledge in a(n) (educational) domain
- We've been at this a while
- Quantified epistemology



Samkhya, Yoga

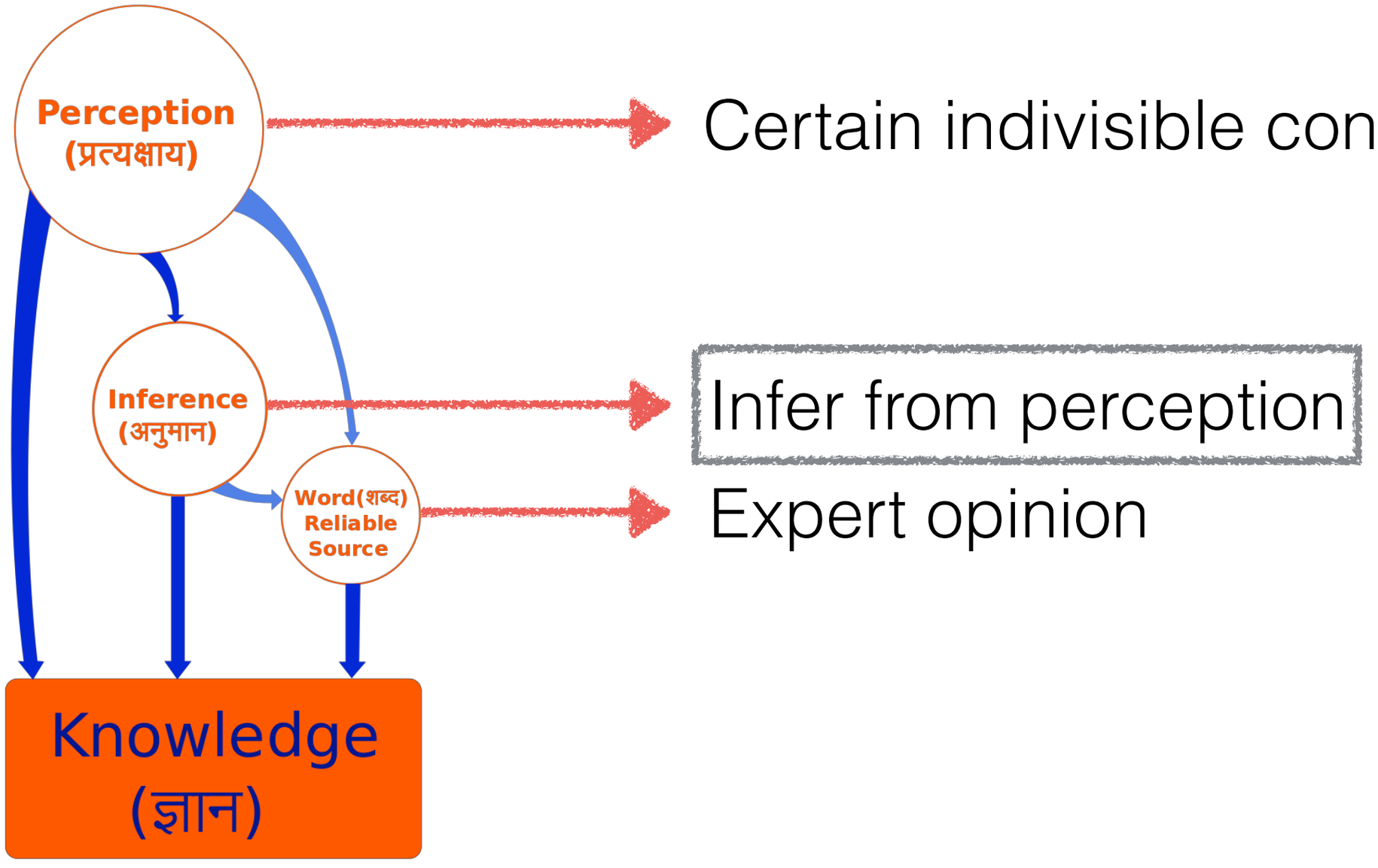
सांख्य
~500BCE



Plato ~300BCE

孟轲 ~ 200BCE

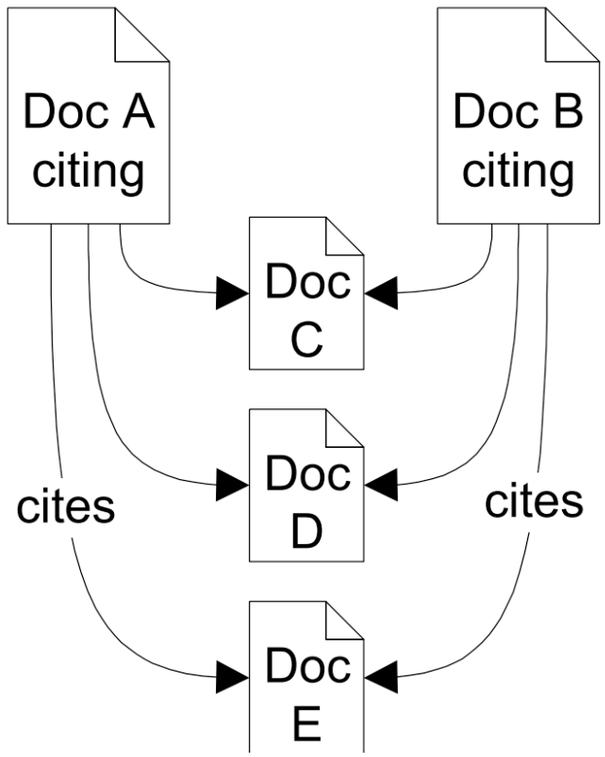
Domain Structure Discovery



Bibliometrics

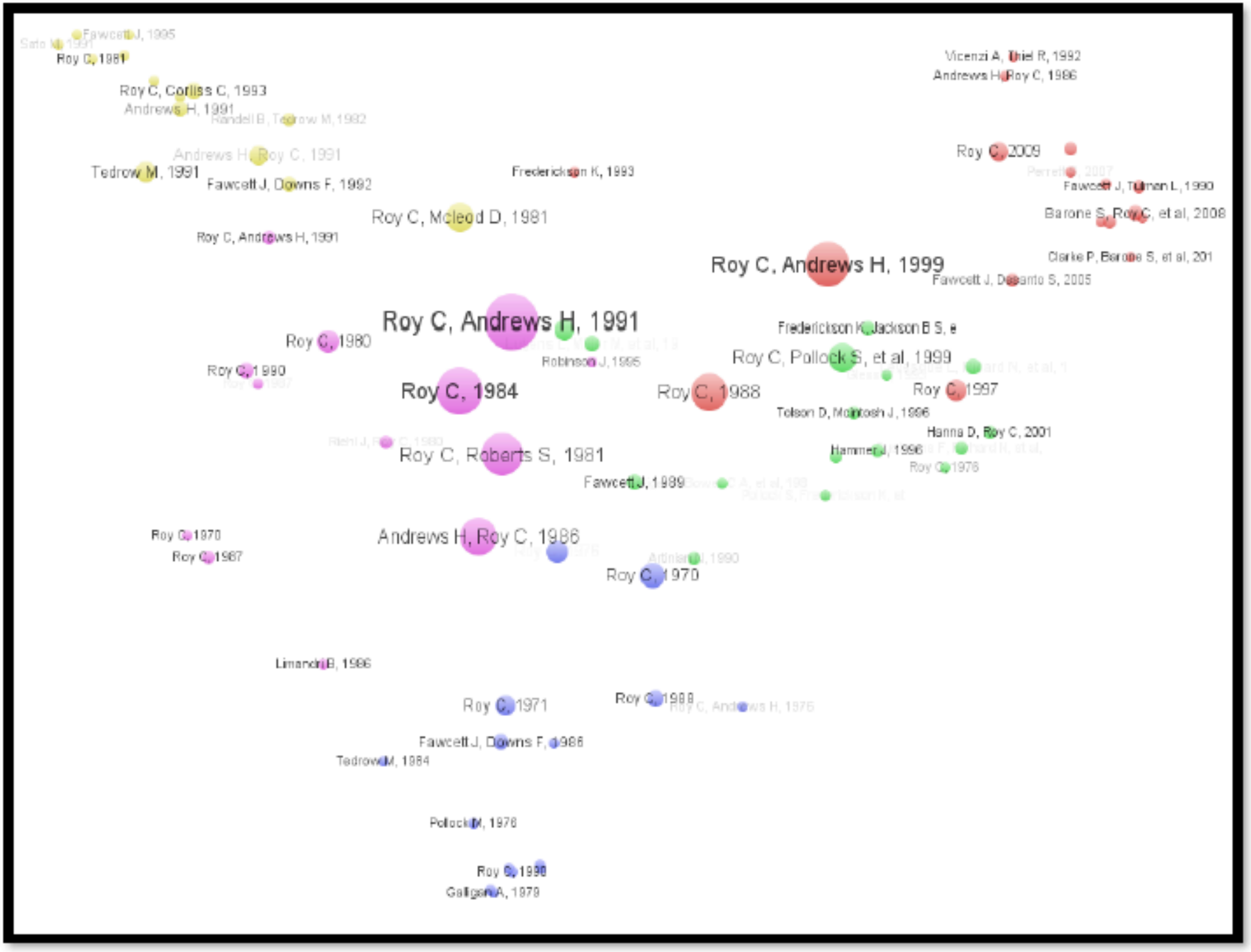
(scientometrics, librametry, statistical bibliography)

- Citation patterns
- Raw number (impact score), Erdős Number
- Co-word analysis
- Network representation

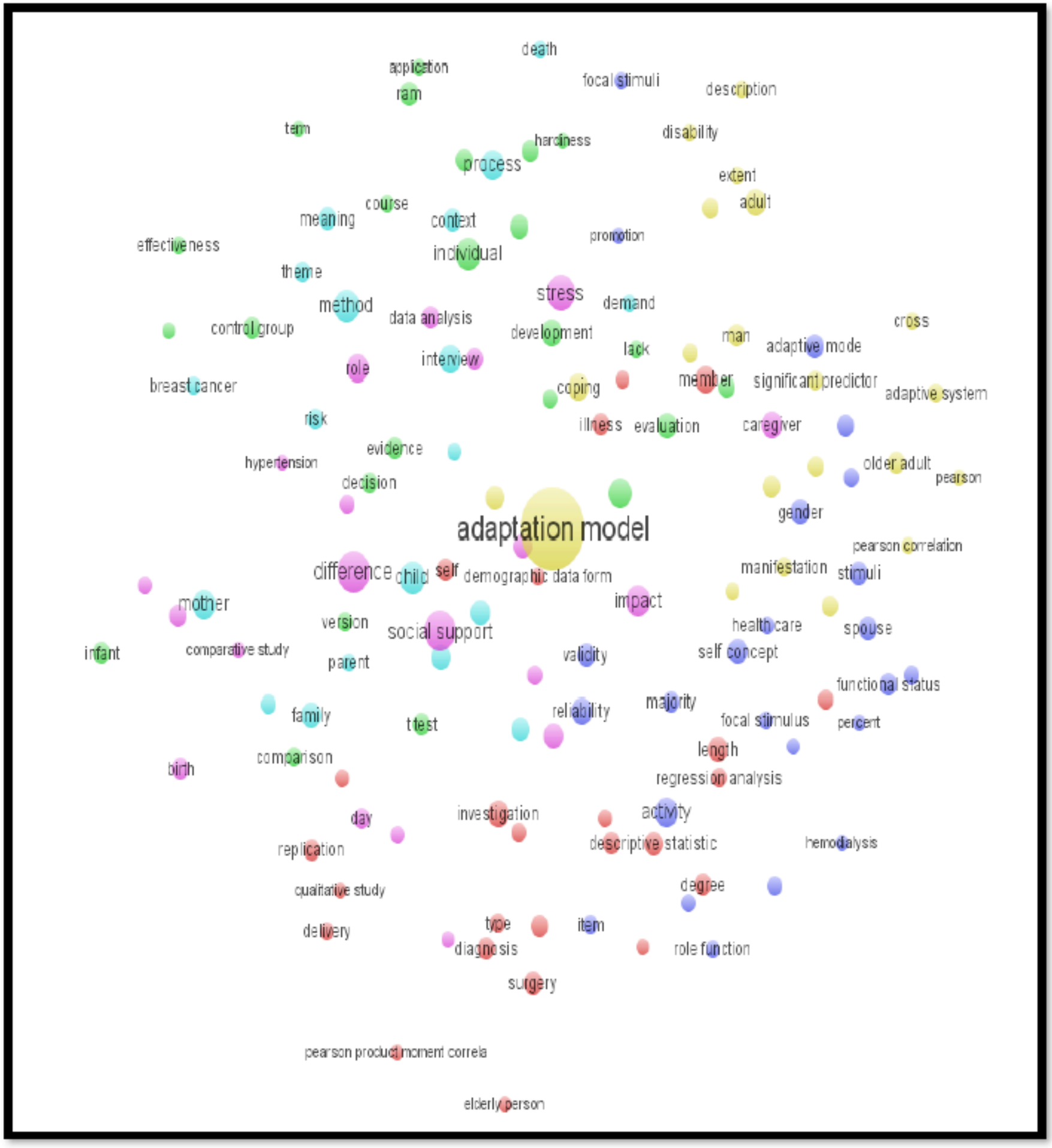


Bibliographic Coupling

(Eigenvectors again!)



Alfonzo, Sakraida, Hastings-Tolsma (2014)



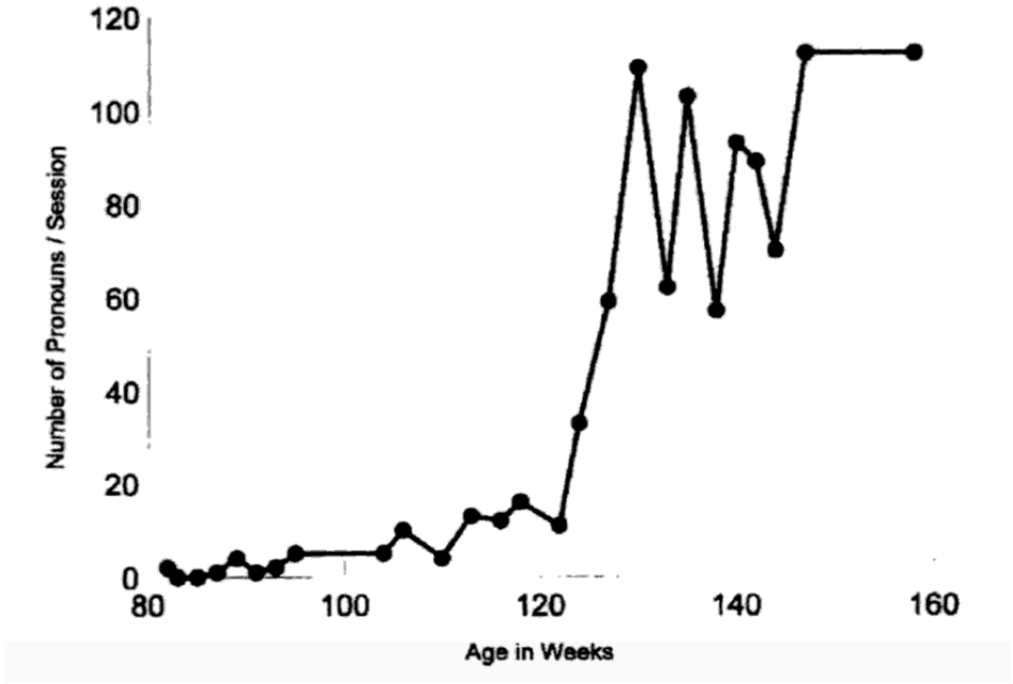
Alfonzo, Sakraida, Hastings-Tolsma (2014)

Latent Variable

Latent variables are variables that are not directly observed but are rather inferred from other variables that are observed and directly measured.

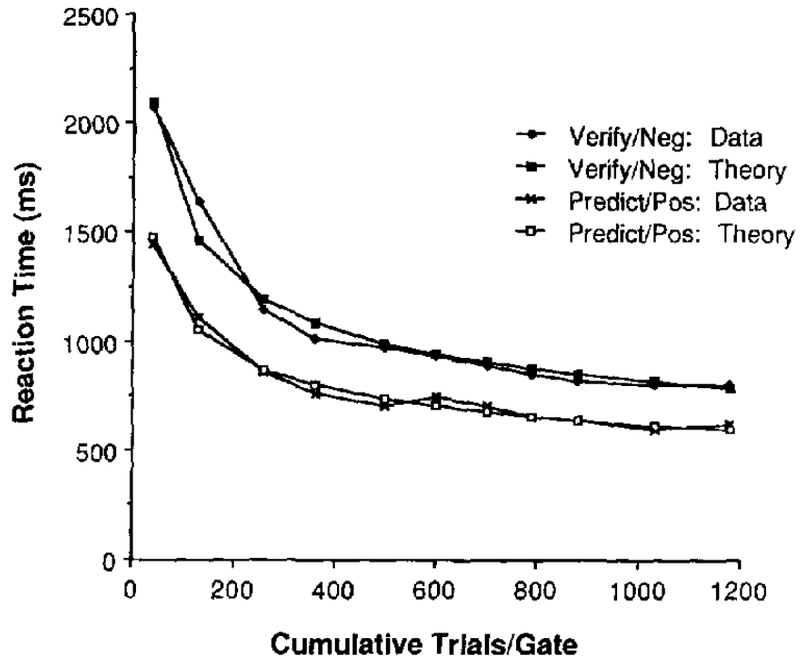
(What isn’t a latent variable?)

Skills



Fischer & Yan, 1980

(There is also the whole world of construct validity)



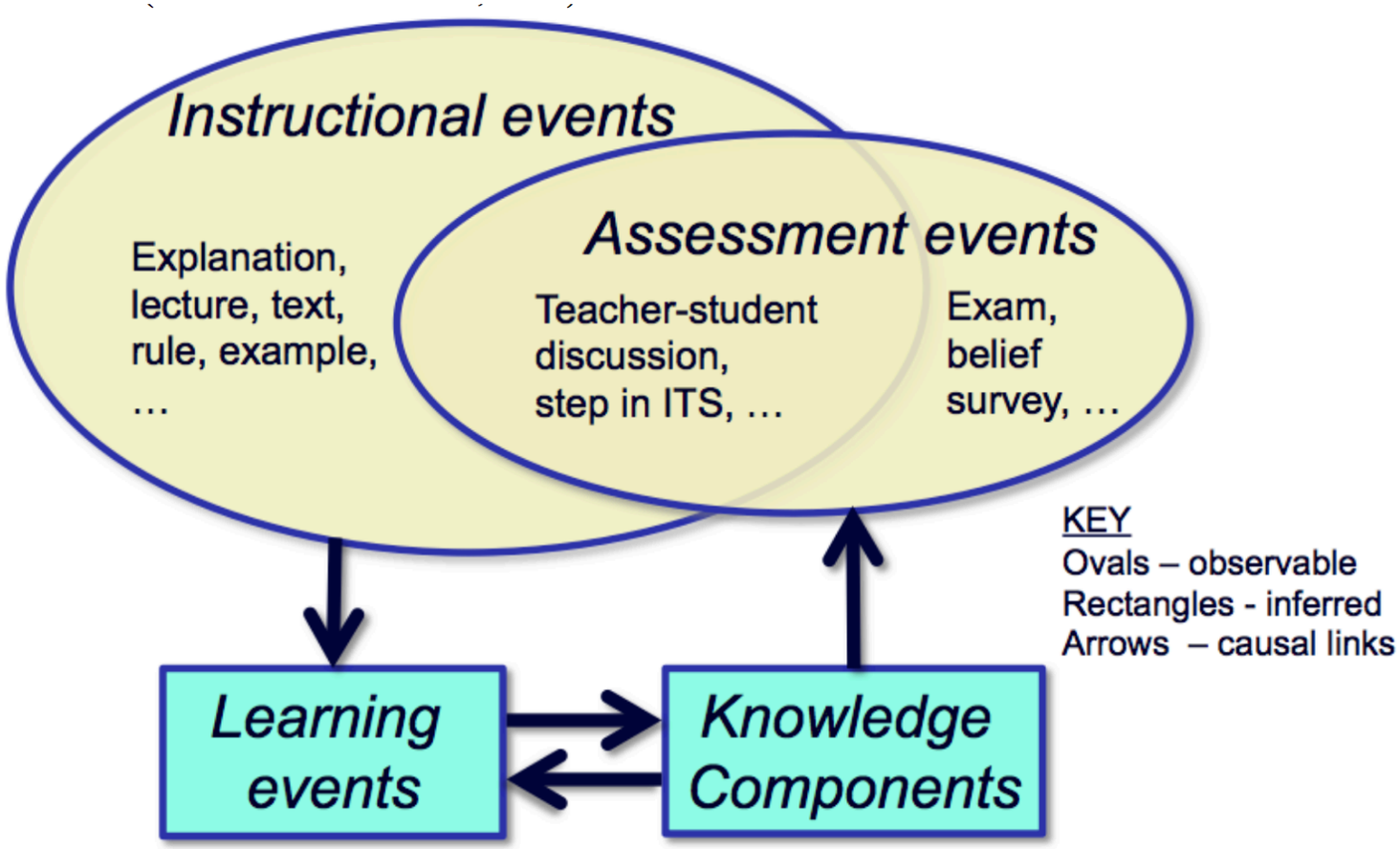
Anderson, 1982

Prolonged learning (memory) about a family of events

Mostly defined by experts/definitionally true

Knowledge Components

A description of a mental structure or process that a learner uses, alone or in combination with other knowledge components, to accomplish steps in a task or a problem (Koedinger & Nathan, 2004)



Q-Matrices

History

- Interested in student misconceptions

- Devised the “Rule Space Method”

- RSM converts item response patterns into probabilities of mastering particular “skills” or concepts

Kikumi Tatsuoka

Q-Matrix

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11	
c1	1	1	1	0	1	0	0	1	0	0	0	
c2	0	0	0	0	1	1	0	0	0	0	0	
c3	1	0	0	0	0	0	0	0	1	1	1	
c4	0	1	0	1	0	1	0	0	0	0	0	
c5	0	0	0	0	0	0	1	1	0	1	0	

Concepts are defined by experts. Very time consuming & domain specific

Q-Matrix

	q1	q2	q3	q4	q5	q6
con1	1	0	0	0	0	1
con2	1	1	0	1	0	0
con3	1	1	1	0	0	0

(Tatsuoka, 1983;1996)

Probability a student is correct given mastery of a given concept

Q-Matrix

	q1	q2	q3	q4	q5	q6
con1	1	0.01	0.6	0	0.7	1
con2	0.8	0.7	0.8	0.76	0.5	0.42
con3	0.5	0.6	1	0.55	0.5	0.67

(Brewer , 1996)

Activity: Build Q-M

- Get into groups of 4
- Agree on a topic
- Agree on 3 concepts within that topic
- Devise 6 questions that relate to the concepts
- Map the concepts to those questions

Activity: Build Q-M

- Now, find another group and have them answer you questions
- Note which ones they get correct/incorrect
- Do the scores map onto your concepts?

Problem

Correspondence between expert-derived Q-matrices and student responses is not 100%

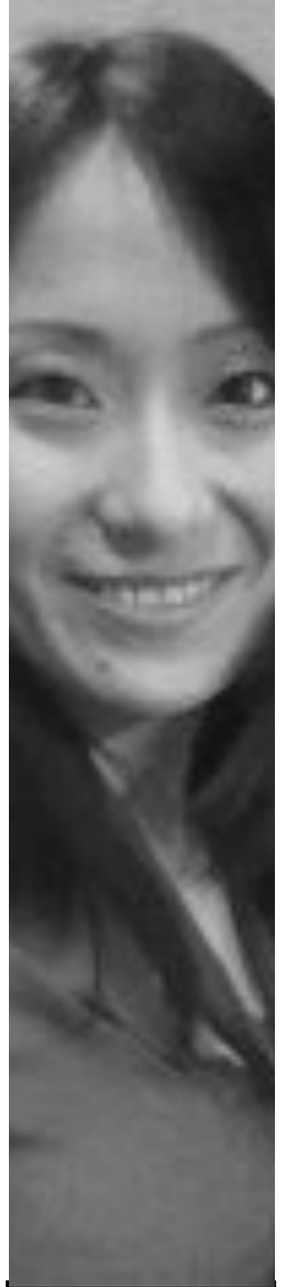
(Hubal, 1992)

Question: Can we use the Q-matrix method to derive valid “student mental states” (constructs? knowledge states? skill definitions?)

Lykken

Bors-boom

de la Torre



Xu

Cron-bach

Fischer

Adkins

Banaji

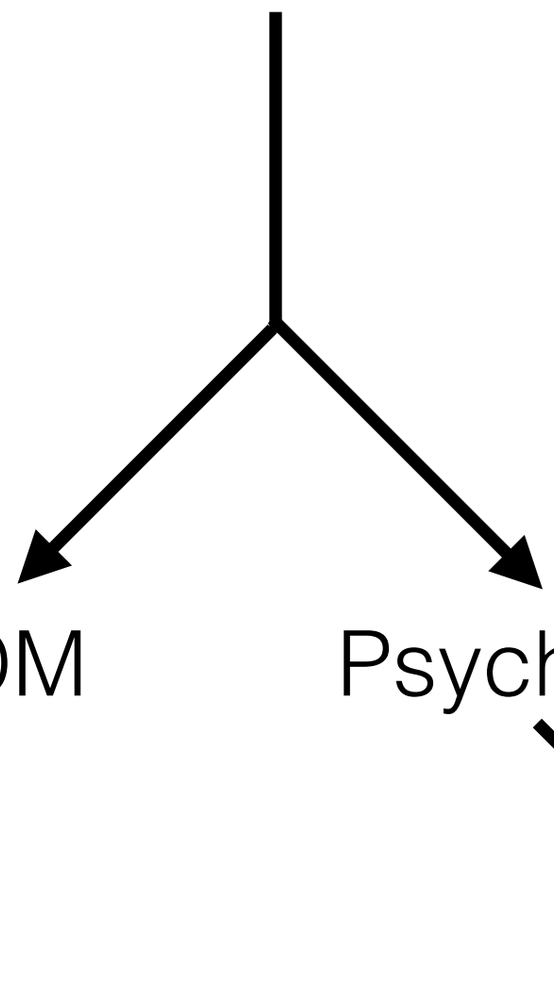
Steele

No

Can this problem be solved?

Yes

Divergence by Domain



Tutoring Systems
Automation
Identifying KS

Large scale assessments
Identifying cognitive states
for use by instructors

One Solution

- Create idealized patterns
- Compare the observed pattern to the idealized
- Use difference between them as an indicator of “model fit”

Idealized Pattern

	q1	q2	q3	q4	q5	q6
c1	1	0	0	0	0	1
c2	1	1	0	1	0	0
c3	1	1	1	0	0	0

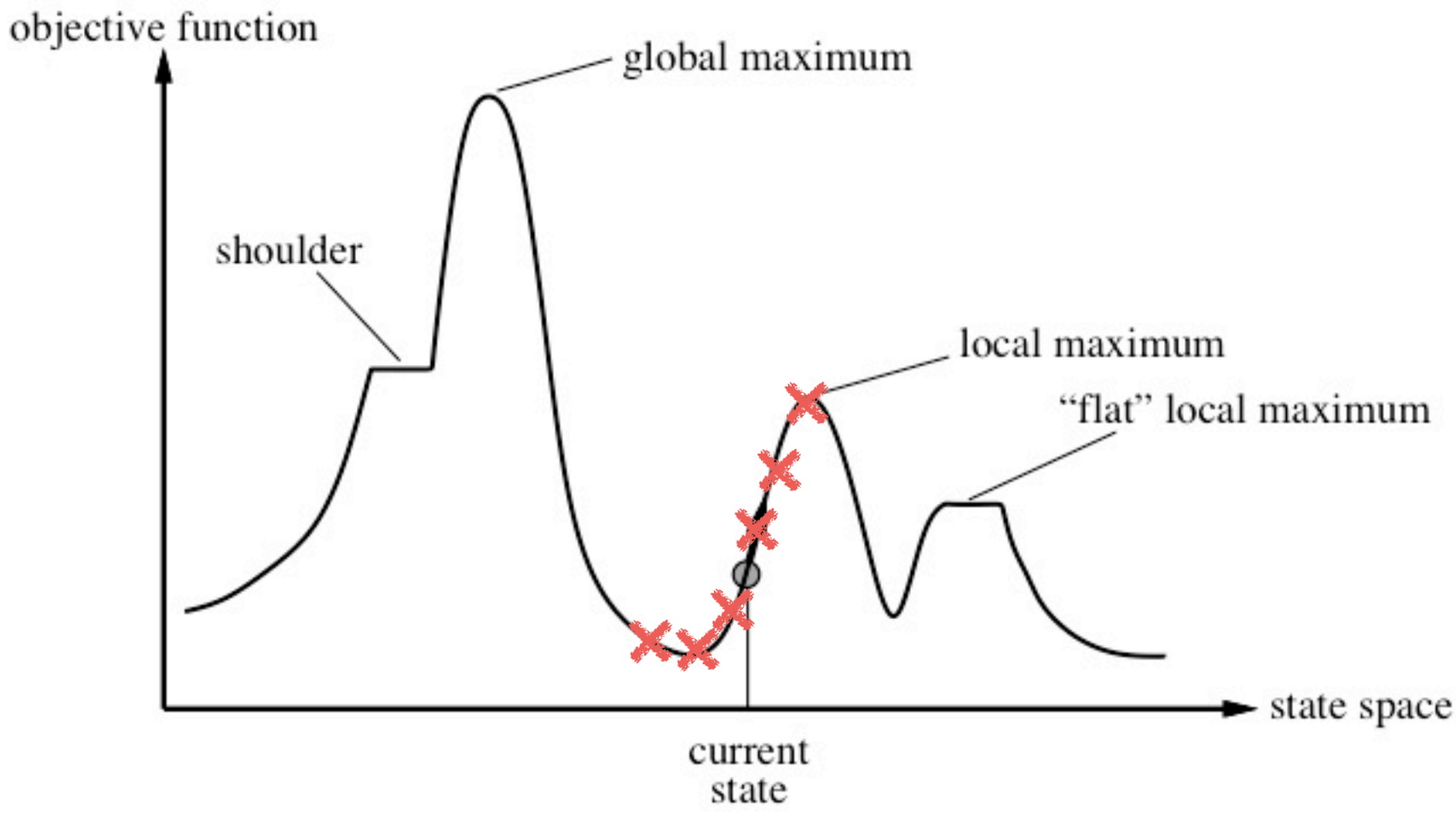
$$L_1 = \quad d(p,IDR) = \sum_q |p(q) - IDR(q)|$$

$$L_1 = \ 1$$

Student Answer: 101110

Concept State	Ideal Response Vector
000	000010
001	001010
010	000110
011	011110
100	000011
101	001011
110	000111
111	111111

Hill Climbing Algorithm



- If we stop too early might only capture a local maxima
- This is a "heuristic" algorithm - when problem is not algebraically solvable or would take too long
- State description contains all the information needed to find a solution