Universitat Pompeu Fabra
December 15, 2016

Data-Driven Design of Study Plans

Rakesh Agrawal
President, Data Insights Laboratories
Rukmini Chair Professor, Indian Institute of Science

Synthesizing Study Plans

Imagine you are an instructor who wants to offer a new course

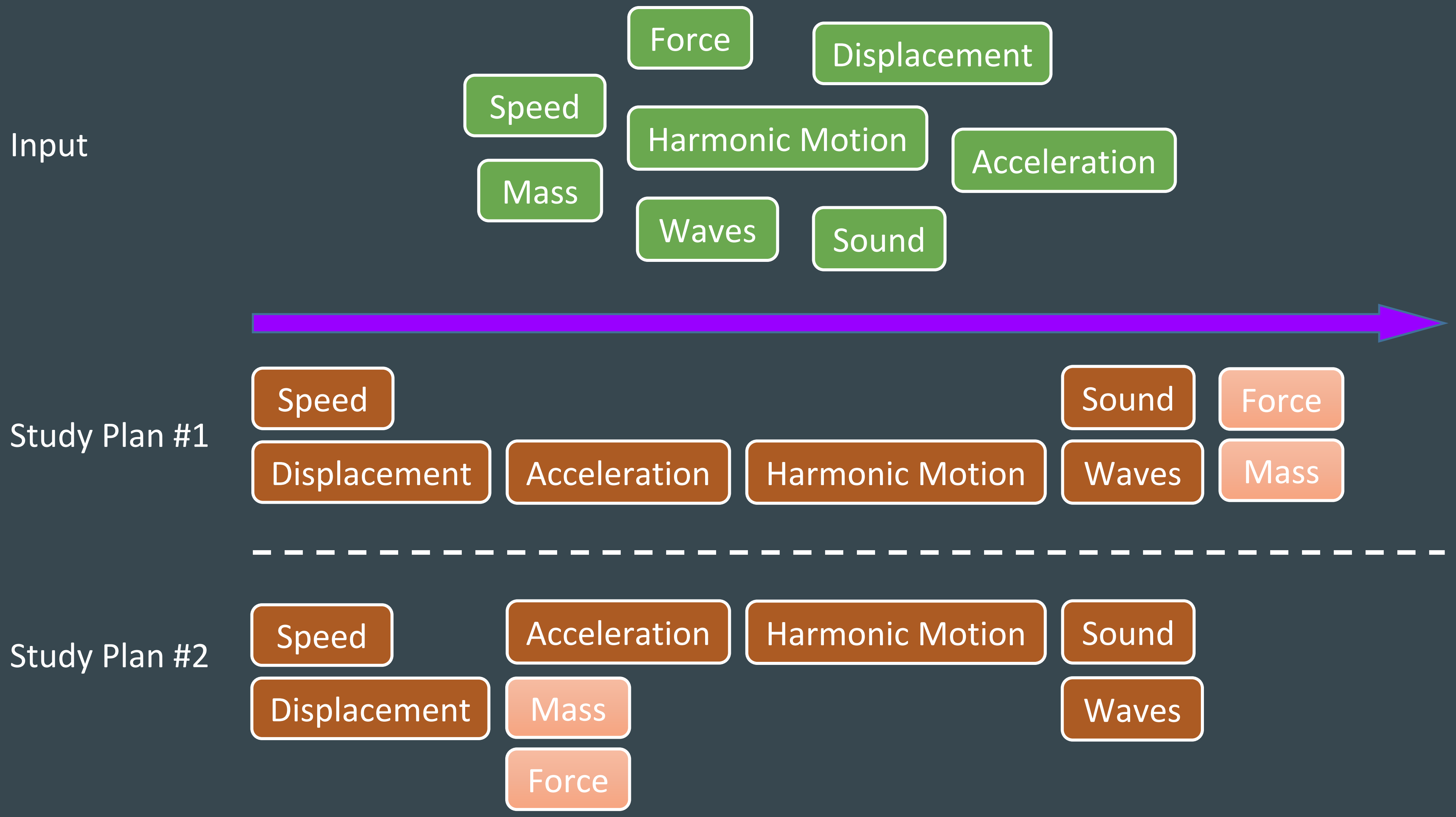
You know the concepts you want to teach in the course, but need help with formulating the study plan:

a. What concepts should you cover in one session

b. The sequencing of sessions

Joint Work with Behzad Golshan and Evangelos Papalexakis [EDM 2016]

Study Plans


Input
Force
Displacement
Speed
Harmonic Motion
Acceleration
Mass
Waves
Sound
Study Plan #1
Speed
Displacement
Acceleration
Harmonic Motion
Sound
Waves
Force
Mass
Study Plan #2
Speed
Displacement
Acceleration
Mass
Force
Harmonic Motion
Sound
Waves

Why is this problem important?

The increasing demand for many more courses of smaller duration on specialized topics

The trend accentuated by increasing availability of inexpensive devices connected to fast Internet

- 250 million smartphones in India within 2 years!

Outline

- Problem Statement
- Related Work
- Our Method
- Experimental Results
- Future Work & Conclusions

Outline

- Problem Statement
- Related Work
- Our Method
- Experimental Results
- Conclusions & Future Work

Axioms

- Learning Unit
  - *A group of coherent concepts suitable to be covered together*
  - Cohesion: Concepts within a learning unit must be closely related
  - Isolation: Concepts in different learning units must be independent
  - Unity: A concept should be covered in one unit
- Study plan
  - *An ordering of some number of learning units*
  - Prerequisite compliance: L1 < L2 => concepts in L2 not needed for L1
  - Locality of references: L2 builds upon L1 => L2 should come soon after L1

Problem Statement

Given a set of concepts,

- Partition them into a given number of learning units, and
- Provide a sequencing of learning units

such that an objective function f is minimized

PROBLEM 1 (STUDY PLAN DESIGN PROBLEM). *Given a concept graph* $G = \langle V, E \rangle$ *with* $n > 0$ *nodes, and the number of desired learning units* m ($m \leq n$), *output an ordered vector of learning units* $\mathcal{L} = \langle L_1, L_2, \cdots, L_m \rangle$ *to*

$$\text{Minimize: } f(\mathcal{L})$$
$$\text{s.t. } \forall i : L_i \subseteq V, L_i \neq \phi, \text{ and}$$
$$\cup_i L_i = V.$$

Objective Function

$$f(\mathcal{L}) = \sum_{\substack{\pi(u)<\pi(v)\\(u,v)\in E}} (\pi(v)-\pi(u)) * C_r,$$

$$+ \sum_{\substack{\pi(u)>\pi(v)\\(u,v)\in E}} (\pi(u)-\pi(v)) * C_p +$$

$$+ \sum_{\substack{\pi(u)=\pi(v)\\(u,v)\notin E}} C_d.$$

Prerequisite Compliance Violation

Locality of Reference Violation

Cohesion Violation (Also Isolation)

Unity Violation Penalized by 1st two terms

Outline

- Problem Statement
- Related Work
- Our Method
- Experimental Results
- Conclusions & Future Work

Related work

On concept maps:

[1] M. Nousiainen, I. T. Koponen. Concept maps representing knowledge of physics: connecting structure and content in the context of electricity and magnetism. *NORDINA* (2010)

[2] I. T. Koponen, M. pehkonen. Coherent knowledge structures of physics represented as concept networks in teacher education. *Science & Education.* Springer (2010)

[3] J. D. Novak, A. J. Canas. The theory underlying concept maps and how to construct and use them. *Technical Report (2006)*

On learning spaces:

[1] J. C. Falmagne, J. P. Doignon. Learning Spaces. *Interdisciplinary Applied Mathematics.* Springer (2011)

[2] A. Y. Kolb, D. A. Kolb. Learning Styles and Learning Spaces: Enhancing Experimental Learning in Higher Education. *Academy of Management learning and education* (2005)

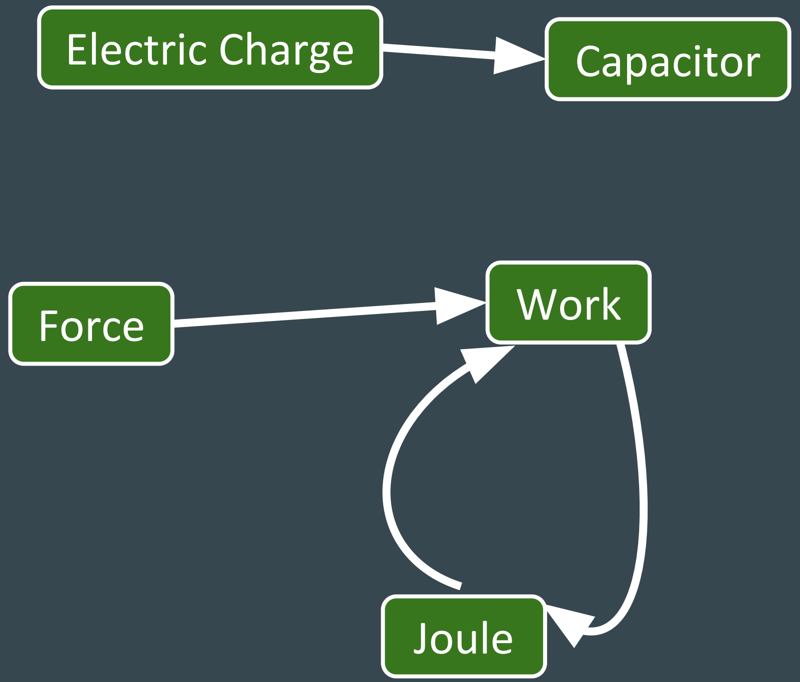
[3] D. Oblinger. Learning Spaces. *EDUCAUSE* (2006)

Outline

- Problem Statement
- Related Work
- Our Method
- Experimental Results
- Conclusions & Future Work

Concept Graph

- Directed graph G = <V,E>
  - Nodes correspond to concepts
  - Edge e = (u, v) implies
    - u and v are related concepts
    - u is a prerequisite of v



Method

0) Creating the Graph

1) Finding Learning Units

2) Ordering Learning Units

3) Organizing Learning Units

Method

On identifying prerequisites:

[1] A. Vuong, T. Nixon, B. Towle. A method for finding prerequisites within a curriculum. *EDM* (2011)

[2] E. Brunskill. Estimating prerequisite structure from noisy data. *EDM* (2011)

Inferring the graph using Wikipedia:

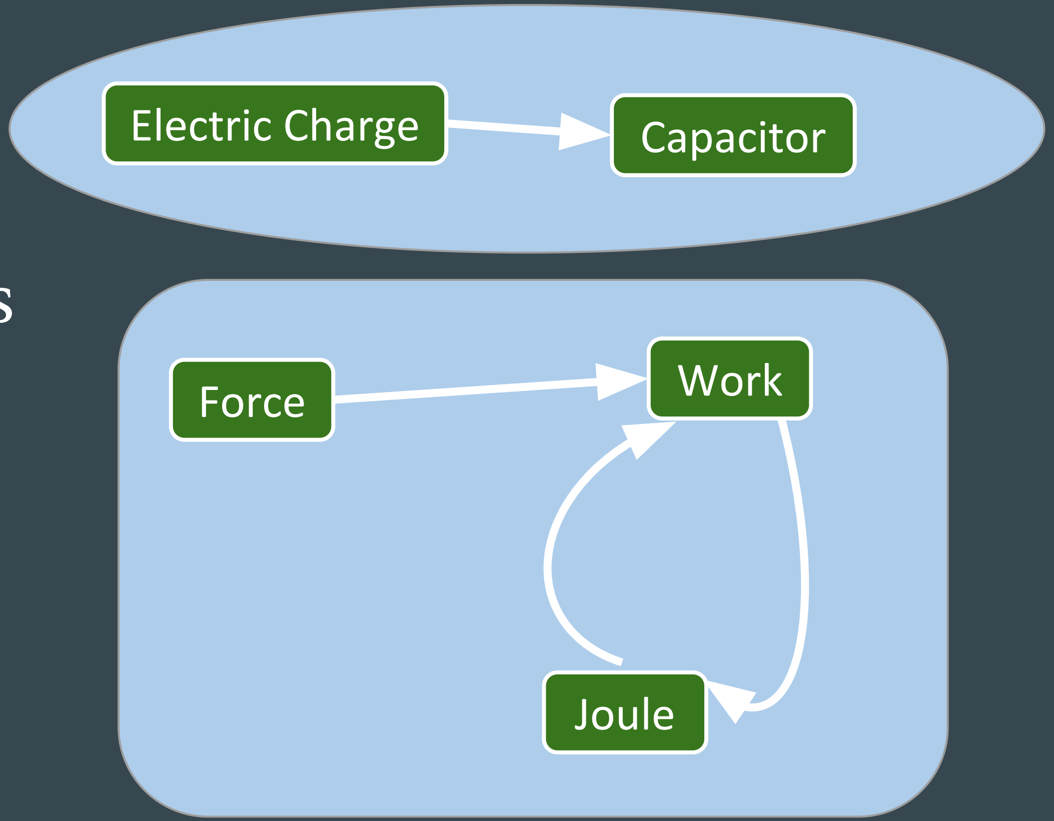
To be discussed along with experimental results

Method

0) Creating the Graph → **1) Finding Learning Units** → 2) Ordering Learning Units → 3) Organizing Learning Units

Partition the graph into dense communities

- Spinglass (Reichardt et al. 2006)
- Allows to adjust the cost of missing edges
- Available in the statistical package R

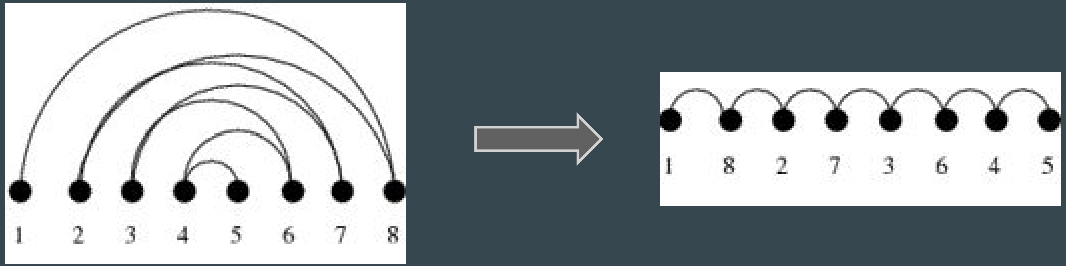


Method

0) Creating the Graph → 1) Finding Learning Units → 2) Ordering Learning Units → 3) Organizing Learning Units

Ordering the learning units

- Shorter forward edges
- Fewer and Shorter backward edges



Theorem: This is NP-hard.

- Proof: Reducing the Minimum Linear Arrangement problem to our problem

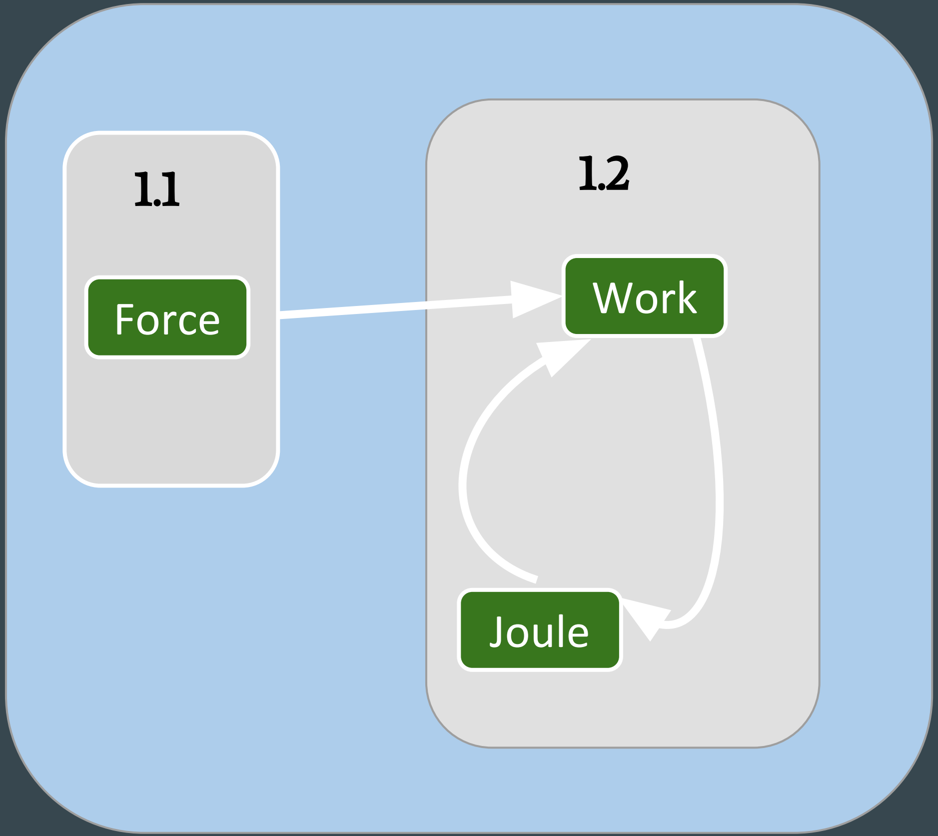
Effective heuristics exist for MinLA problem

- Simulated annealing [SS98]

Method

0) Creating the Graph → 1) Finding Learning Units → 2) Ordering Learning Units → **3) Organizing Learning Units**

- Step 1: Find the Strongly Connected Components
- Step 2: Sort topologically the components
  - Not unique (This implies flexibility)



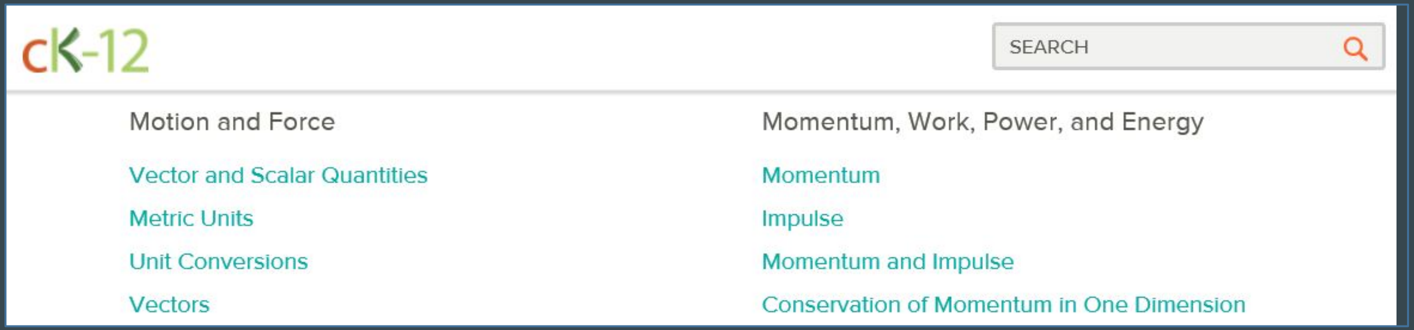
Outline

- Problem Statement
- Related Work
- Our Method
- Experimental Results
  - Creating the Concept Graph
  - Synthesizing the Study Plan
- Conclusions & Future Work

Concept Graph

0) Creating the Graph → 1) Finding Learning Units → 2) Ordering Learning Units → 3) Organizing Learning Units

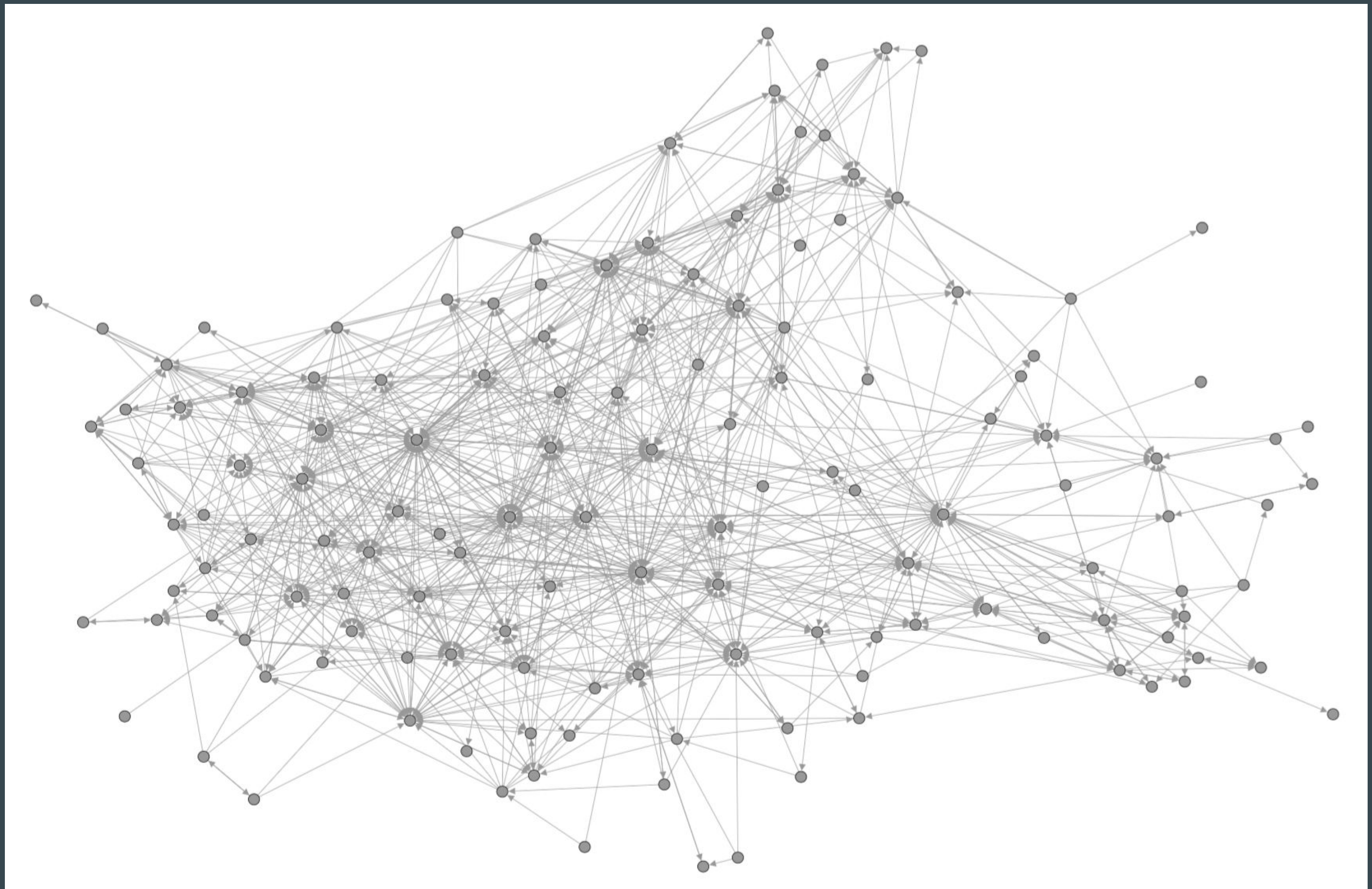
Input: 139 high school physics concepts from CK12.org



Question: Can we use Wikipedia to induce the concept graph?

- Map each concept to a Wikipedia page
- Find the Wikipedia graph between these pages
  - Edges are hyperlinks that link to another page

The CK12 Wikipedia Graph



Problem in using Wikipedia Graph as Concept Graph

Edges do not capture the prerequisite relationships

Two main sources of errors:

- Informative edge, but the direction is wrong

  "capacitors" → "Joule"

  > Conventional capacitors provide less than 360 joules per kilogram of energy

- Unrelated edges

  "capacitors" → "lasers"

  > low-inductance high-voltage capacitors (*capacitor banks*) are used to supply huge pulses of current for many pulsed power applications. These include electromagnetic forming, Marx generators, pulsed lasers

Correcting Errors

- A machine learning approach
  - We manually labeled the edges.
- Features
  - In-degree
  - Out-degree
  - # of languages
  - # of categories
  - …

Many misclassified as unrelated

Actual Edge Type	Prediction Outcome: prereq.	unrelated	reverse
prereq.	29	142	1
unrelated	9	668	8
reverse	0	110	42

But few are misclassified in the wrong directions

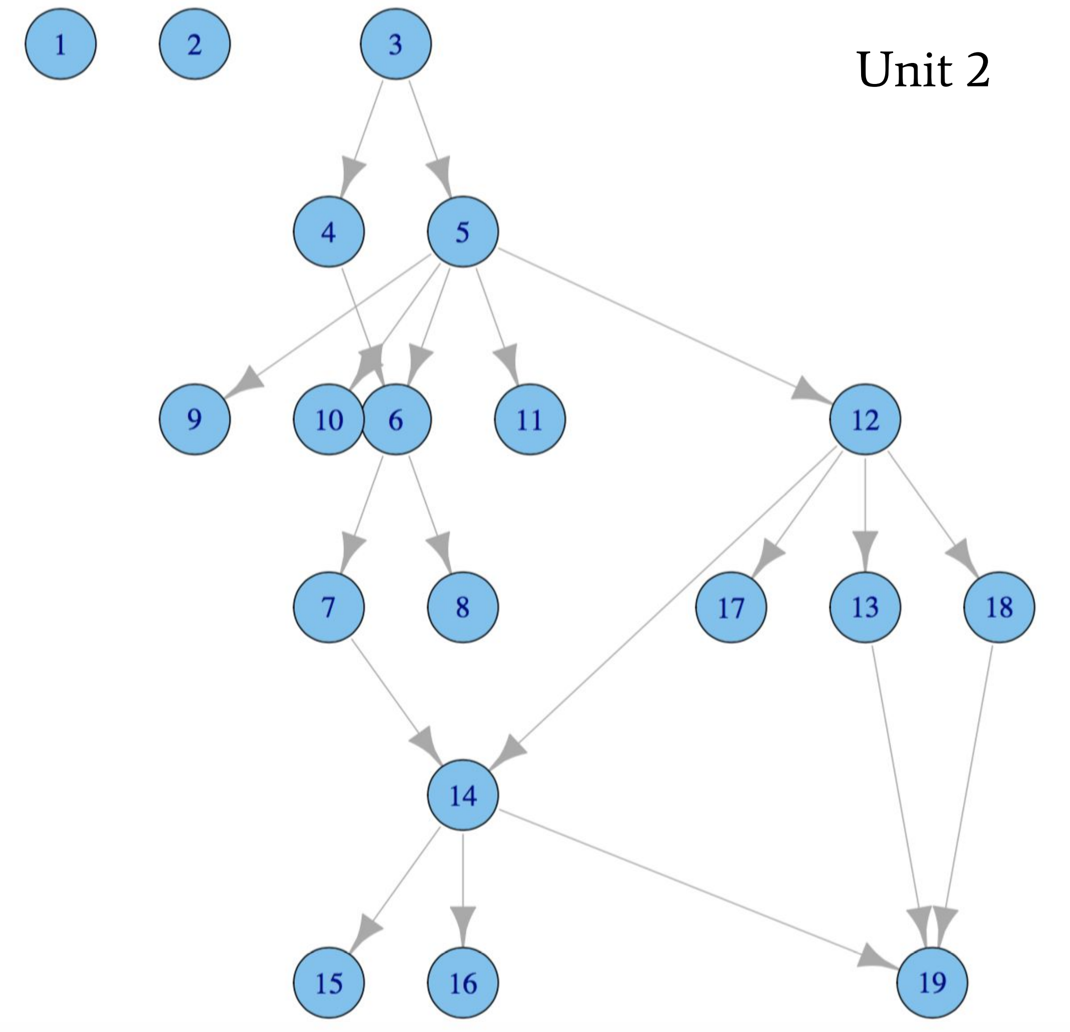
Synthesized Study Plan

Unit 1(20 concepts)	Unit 2 (21 concepts)	Unit 3 (14 concepts)	Unit 4 (18 concepts)
buoyancy	acceleration	atom	calorimetry
euclidean vector	angular momentum	bohr model	change of state
force	angular velocity	conservation of energy	combined gas law
free body diagram	centripetal force	elastic collision	conversion of units
friction	circular motion	inelastic collision	double-slit experiment
impulse	displacement	kinetic energy	energy
inclined plane	keplers laws of ...	mass versus weight	energy conversion

Unit 5 (27 concepts)	Unit 6 (28 concepts)	Unit 7 (11 concepts)
ammeter	beat	doppler effect
capacitor	color	general relativity
capacitors in circuits	concave lens	half-life
electric charge	conduction	length contraction
electric current	curved mirror	mathematical physics
electric field	diffraction	newtons law of univer...
electric power	diffraction grating	rc time constant

Organizing Learning Units

ID	Concepts	ID	Concepts
1	weightlessness	11	simple harmonic motion
2	projectile motion	12	acceleration
3	motion	13	centripetal force
4	moment of inertia	14	newton's laws of motion
5	vector addition	15	kepler's laws of ...
	displacement	16	orbital motion
	velocity	17	motion graphs
6	angular velocity	18	uniform acceleration
7	angular momentum	19	circular motion
8	rotation around a ...		
9	kinematics		
10	pendulum		



User Study

- Recruited 9 domain experts (Physics teachers, Graduate students)
- They were given the following tasks:
  1) *Count the number of odd concepts in each learning unit that you believe do not belong to the unit*
  2) *Without changing any of the learning units proposed, what order do you suggest?*

Results of the User Study

- Number of concepts that do not belong in the respective unit:

	Min	Max	Median	Mean	# Concepts
Unit 1	1	6	3.0	3.4	20
Unit 2	0	3	1.0	1.1	20
Unit 3	1	7	3.5	3.7	14
Unit 4	0	5	2.0	1.8	18
Unit 5	0	4	1.0	1.0	26
Unit 6	0	3	0.5	0.9	28
Unit 7	0	5	1.0	1.4	11

- Only two participants ordered the units somewhat differently
- The high school Physics teacher: our study plan was very clever

Outline

- Problem Statement
- Related Work
- Our Method
- Experimental Results
- Conclusions & Future Work

Recap

- We formalized the problem of synthesizing study plans automatically
- We provided a novel and pragmatic solution
- Our method did not use domain specific knowledge
  - Generalizing to other areas seems promising
- Our experimental results as well as the user study show that the problem of creating study plans is amenable to computational approaches

Further Work

- Incorporating user modeling into the system
  - Creating study plans that suit students background/interests/abilities
- Investigating how human input (implicit or explicit) can improve the quality of generated study plans

Bigger Picture: Datafication of Education

How to enhance the quality of the electronic textbooks?

How to form teams of students in a class?

How to create study plans for courses?

Data Mining for Enhancing Electronic Textbooks

Diagnostic tools for identifying weaknesses in textbooks

Within section deficiencies

Syntactic complexity of writing and dispersion of key concepts in the section [AGK+11a]

Across sections deficiencies

Comprehension burden due to non-sequential presentation of concepts [ACG+12]

Algorithmic enhancement of textbooks for enriching reading experience

References to selective web content

Links to authoritative articles [AGK+10], images [AGK+11b] and videos [ACG+14] based on the focus of the section

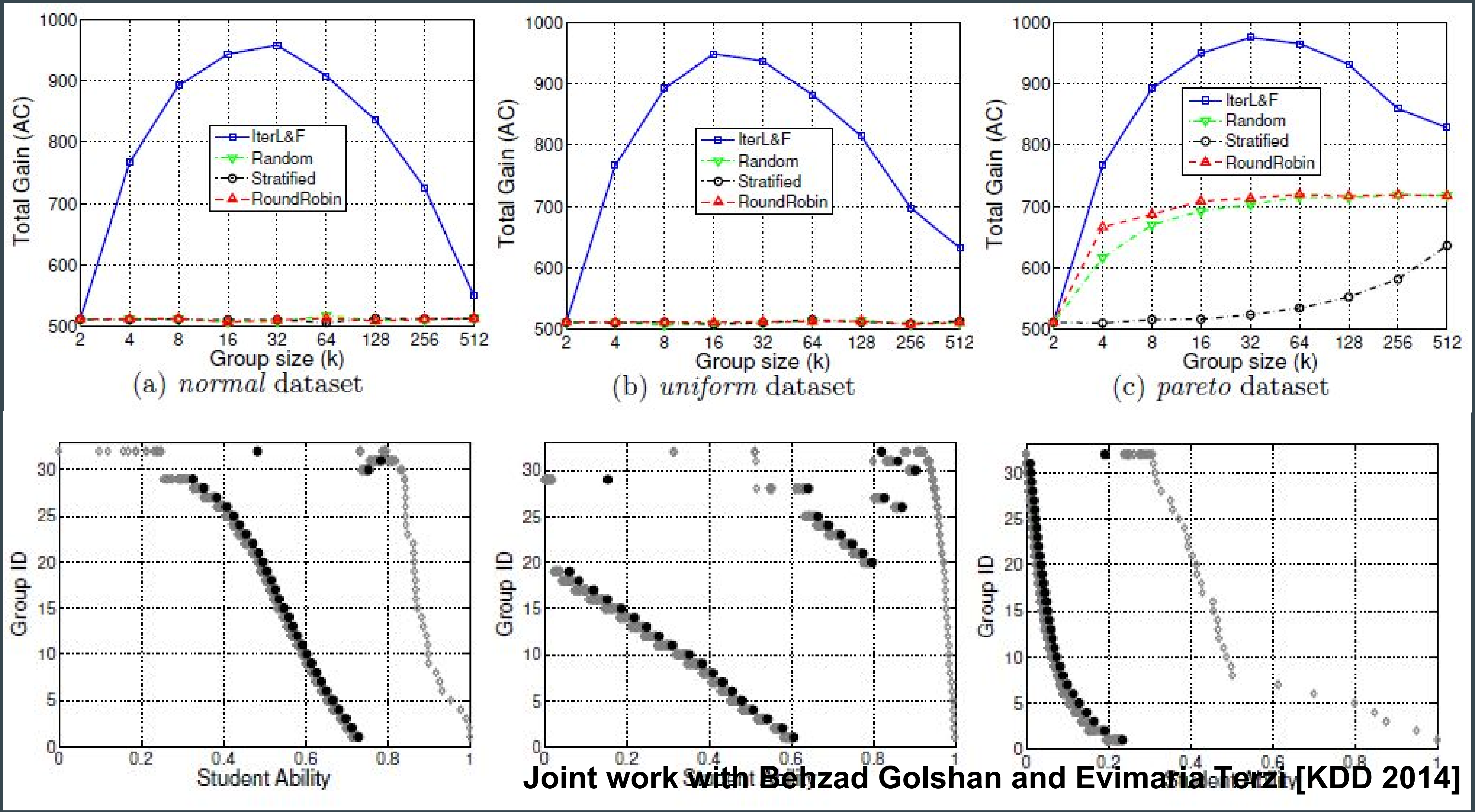
References to prerequisites

Links to concepts necessary for understanding the present section, derived using a model of a how students read textbooks [AGK+13]

Validation on textbooks from U.S.A and India, on different subjects, across grades

Joint work with Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, *et al.*

Forming Beneficial Teams of Students



Joint work with Behzad Golshan and Evimaria Terzi [KDD 2014]

Research Opportunities in Data-Driven Education

- ❖ Validation of experimental results through deployment
- ❖ Synergies with crowd-sourcing approaches
- ❖ Use of logs of interactions data and personalization
- ❖ Performance evaluation methodologies and benchmarks
- ❖ Issues related to privacy, security, confidentiality, copyright, royalty …

Magic happens when what is desperately needed meets what is technically feasible

Selected References

[AGK+10] Rakesh Agrawal, Sreenivas Gollapudi, Krishnaram Kenthapadi, Nitish Srivastava, Raja Velu. "Enriching Textbooks Through Data Mining". DEV 2010.

[AGK+11a] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "Identifying Enrichment Candidates in Textbooks". WWW 2011.

[AGK+11b] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "Enriching Textbooks With Images". CIKM 2011.

[ACG+12] Rakesh Agrawal, Sunandan Chakraborty, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "Empowering Authors to Diagnose Comprehension Burden in Textbooks". KDD 2012.

[AGK+13] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "Studying from Electronic Textbooks". CIKM 2013.

[AJK14] Rakesh Agrawal, M. Hanif Jhaveri, and Krishnaram Kenthapadi. "Evaluating Educational Interventions at Scale". LAS 2014.

[ACG+14] Rakesh Agrawal, Maria Christoforaki, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, Adith Swaminathan. "Augmenting Textbooks with Videos". ICFCA 2014.

[AGT14] Rakesh Agrawal, Behzad Golshan, Evimaria Terzi. "Grouping Students in Educational Settings". KDD 2014.

[AGP16] Rakesh Agrawal, Behzad Golshan, Evangelos Papalexakis. "Toward Data-Driven Design of Educational Courses". EDM 2016.


Questions?