Analytic Combinatorics for the masses

Robert Sedgewick
Princeton University

This talk is dedicated to the memory of Philippe Flajolet

Philippe Flajolet 1948–2011

Starting point

1975: What are the algorithms that *everyone* should know?

"Everyone" means "everyone using a computer"

- scientists
- engineers
- mathematicians
- software/hardware designers
- cryptananalysts
- COBOL programmers
- . . .

Context

- <1 computer per university, on average
- <10 CS departments worldwide

1981: First edition of *Algorithms* is published

Disseminating knowledge: RS context

Algorithms

Analytic Combinatorics

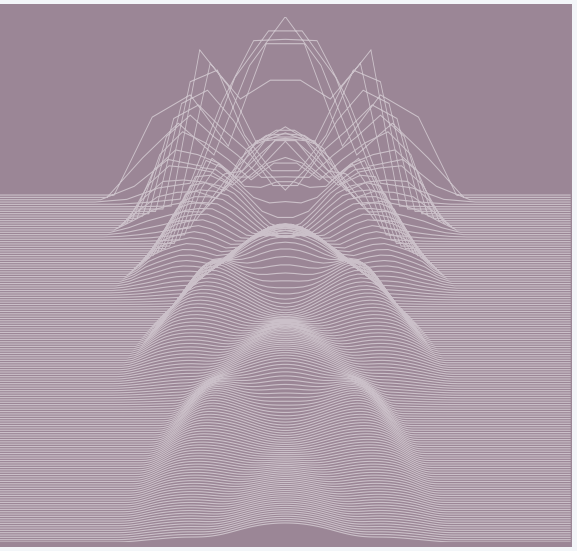
Introduction to Computer Science

1980

1990

2000

2010



Analytic Combinatorics for the masses

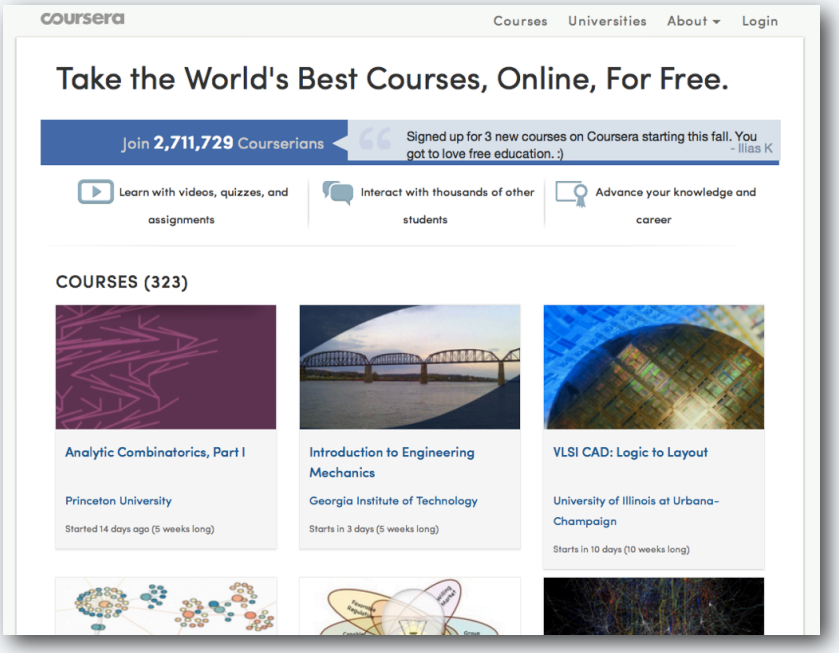
- Frequently asked questions
- Disruptive changes
- A way forward
- Analytic combinatorics
- Questions answered

Frequently asked questions

Q. What does MOOC mean?

A. *Massive Open Online Course.*

- *Free*
- Designed for large numbers of "students"



Just one platform (2013):

300+ courses

2.7 million students

Q. What is the business model?

Q. How are our students getting their money's worth if it's free?

Q. Can we afford to participate?

Q. Will this detract from our "brand"?

Q. Can I get credit for this course?

the wrong questions

Note: A more disruptive "connectivist" MOOC concept is widely advocated:

- Crowdsourced decisions and assessment
- Aggregation, remixing, and passing forward content

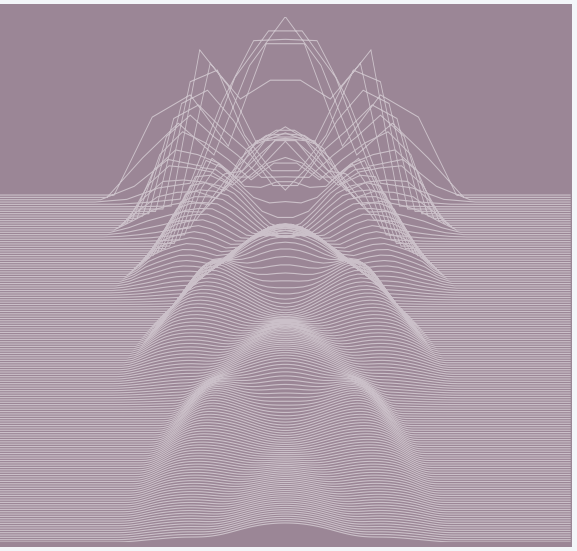
The right questions

Q. How are we going to disseminate knowledge in the future?

Q. Are universities going to take leadership in helping to decide this question?

Q. If we *can* disseminate knowledge for free, isn't it our responsibility to do so?

Q. What is the purpose of a university?



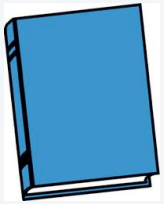
Analytic Combinatorics for the masses

- Frequently asked questions
- **Disruptive changes**
- A way forward
- Analytic combinatorics
- Questions answered

Seismic changes are afoot

Records? 

Newspapers? 

Books? 

Libraries? 

Universities?

For a millennium, universities have been considered the main societal hub for knowledge and learning. And for a millennium, the basic structures of how universities produce and disseminate knowledge and evaluate students have survived intact...Today, though, the business of higher education seems to some as susceptible to tech disruption as other information-centric industries

The Future of Higher Education - Pew Internet & American Life Project, 2012

Business of higher education ?? A road to ruin.

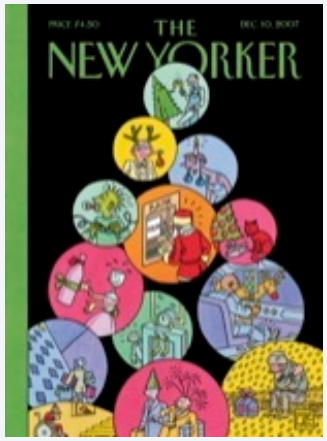
Grafton: Save the libraries

FUTURE READING
Digitization and its discontents.
by Anthony Grafton
The New Yorker
November 5, 2007

. . .Sit in your local coffee shop, and your laptop can tell you a lot. If you want deeper, more local knowledge, you will have to take the narrower path that leads between the lions and up the stairs. There—as in great libraries around the world—you'll use all the new sources, the library's and those it buys from others, all the time. You'll check musicians' names and dates at Grove Music Online, read Marlowe's "Doctor Faustus" on Early English Books Online, or decipher Civil War documents on Valley of the Shadow. But these streams of data, rich as they are, will illuminate, rather than eliminate, books and prints and manuscripts that only the library can put in front of you. ***The narrow path still leads, as it must, to crowded public rooms where the sunlight gleams on varnished tables, and knowledge is embodied in millions of dusty, crumbling, smelly, irreplaceable documents and books.***

RS: Think about the future

The New Yorker
Letter to the editor
Robert Sedgewick
December 10, 2007

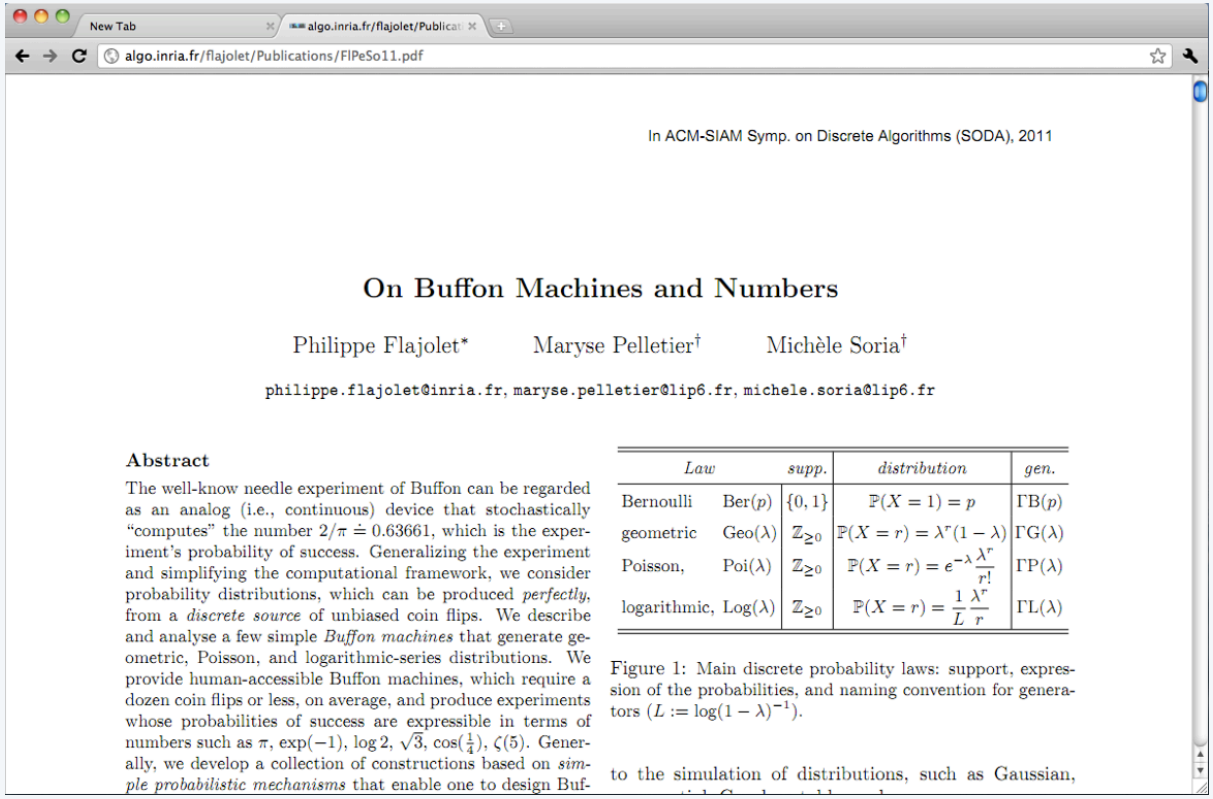


While Grafton's reservations about putting knowledge online are well taken, I would also point out that there is quite a bit going on now in the academic world that doesn't have much to do with old books. Indeed, as the author of many books, I wonder whether perhaps the book is not quite sacred as a means of disseminating knowledge. ***What is the most effective way to produce and disseminate knowledge with today's technology? How can we best structure what we know and learn so that students, researchers, and scholars of the future can best understand the work of today's researchers and scholars?*** *I think that questions like these are more important and more difficult to address than whether we can put the contents of libraries on the Web.*

Disseminating knowledge I: research papers

When is the last time you visited a library to find a paper?

Did you print the papers to read the last time you refereed a conference?



New Tab | algo.inria.fr/flajolet/Publicat

algo.inria.fr/flajolet/Publications/FlPeSo11.pdf

In ACM-SIAM Symp. on Discrete Algorithms (SODA), 2011

On Buffon Machines and Numbers

Philippe Flajolet* Maryse Pelletier† Michèle Soria†

philippe.flajolet@inria.fr, maryse.pelletier@lip6.fr, michele.soria@lip6.fr

Abstract

The well-know needle experiment of Buffon can be regarded as an analog (i.e., continuous) device that stochastically "computes" the number $2/\pi \doteq 0.63661$, which is the experiment's probability of success. Generalizing the experiment and simplifying the computational framework, we consider probability distributions, which can be produced *perfectly*, from a *discrete source* of unbiased coin flips. We describe and analyse a few simple *Buffon machines* that generate geometric, Poisson, and logarithmic-series distributions. We provide human-accessible Buffon machines, which require a dozen coin flips or less, on average, and produce experiments whose probabilities of success are expressible in terms of numbers such as π, $\exp(-1)$, $\log 2$, $\sqrt{3}$, $\cos(\frac{1}{4})$, $\zeta(5)$. Generally, we develop a collection of constructions based on *simple probabilistic mechanisms* that enable one to design Buf-

Law		*supp.*	*distribution*	*gen.*
Bernoulli	$\mathrm{Ber}(p)$	$\{0,1\}$	$\mathbb{P}(X=1)=p$	$\Gamma\mathrm{B}(p)$
geometric	$\mathrm{Geo}(\lambda)$	$\mathbb{Z}_{\geq 0}$	$\mathbb{P}(X=r)=\lambda^r(1-\lambda)$	$\Gamma\mathrm{G}(\lambda)$
Poisson,	$\mathrm{Poi}(\lambda)$	$\mathbb{Z}_{\geq 0}$	$\mathbb{P}(X=r)=e^{-\lambda}\frac{\lambda^r}{r!}$	$\Gamma\mathrm{P}(\lambda)$
logarithmic,	$\mathrm{Log}(\lambda)$	$\mathbb{Z}_{\geq 0}$	$\mathbb{P}(X=r)=\frac{1}{L}\frac{\lambda^r}{r}$	$\Gamma\mathrm{L}(\lambda)$

Figure 1: Main discrete probability laws: support, expression of the probabilities, and naming convention for generators ($L := \log(1-\lambda)^{-1}$).

to the simulation of distributions, such as Gaussian,

- **Color?**
- **Links to references?**
- **Links to detailed proofs?**
- **Simulations?**

why not?

"I could read it on my iPad ...if I had an iPad"
***D. E. Knuth* (2011)**

why?

Question: If it will not be read on paper, why write it as if it will?

Prediction: Someone will soon invent the future (should be easy)

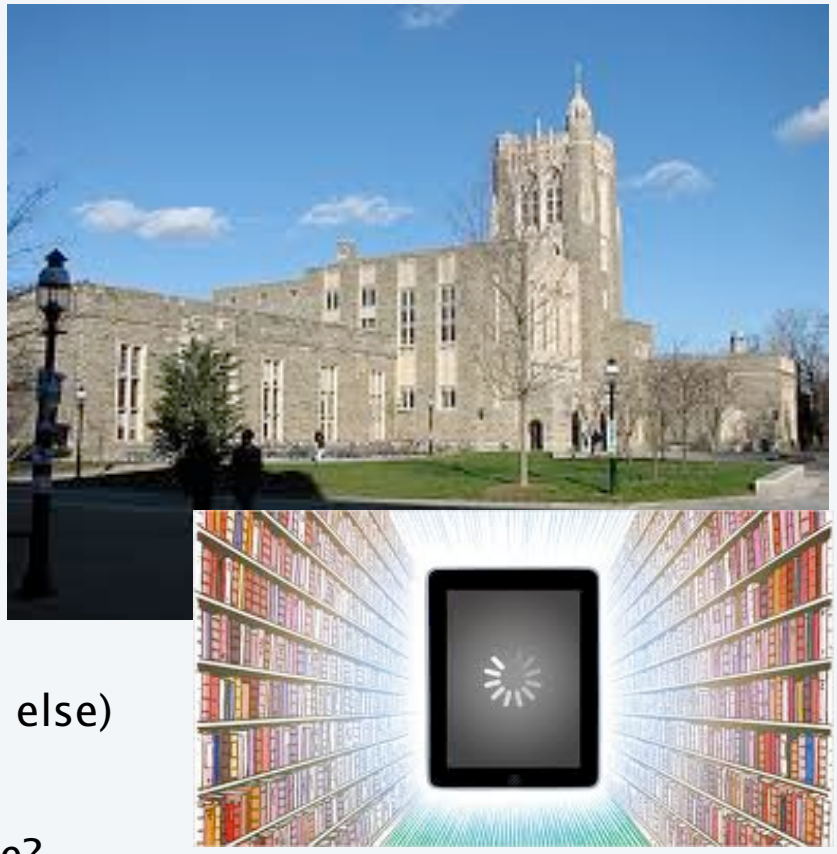
Future of libraries?

1980s

- Students spend significant time in the library
- Faculty members depend on the library for research

2010s

- Students spend significant time online and have *no need* for the library
- Few faculty members in the sciences use the library *at all* for research



2020s?

- A few *book museums* (for Grafton)
- Digital library infrastructure (for everyone else)

How will we disseminate knowledge in the future?

Will universities play a role?

Disseminating knowledge II: Textbooks

We are on a road to ruin

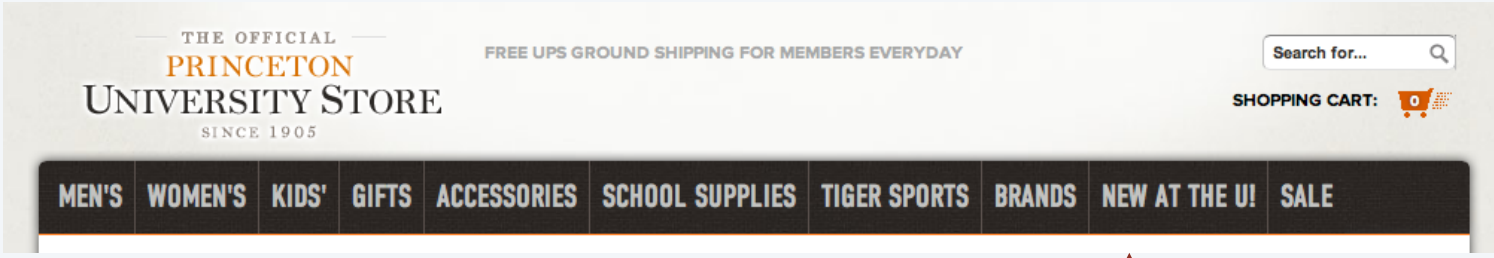
- Prices continue to escalate.
- Students now *rent*, not own books.
- Planned obsolescence? Walled gardens?

$89.4 million
Rise in dollar sales of college textbooks, 2004-5

600,000
Decline in number of new college textbooks sold, 2004-5

Princeton U-store 1950s

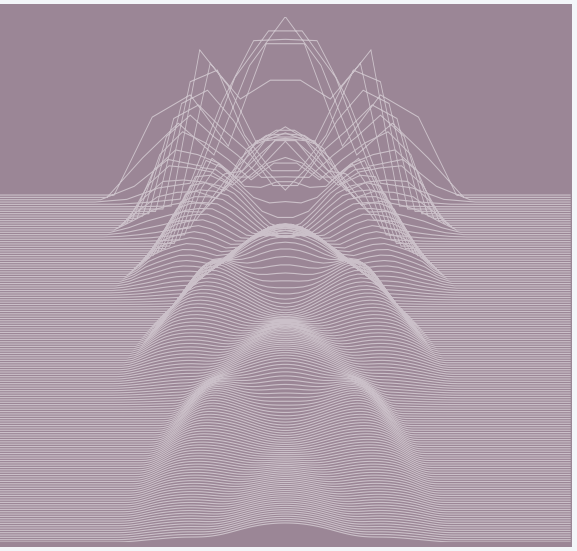
Princeton U-store 2010s



Is there room for a good textbook?

Will free web resources prevail?

No books!



Analytic Combinatorics for the masses

- Frequently asked questions
- Disruptive changes
- A way forward
- Analytic combinatorics
- Questions answered

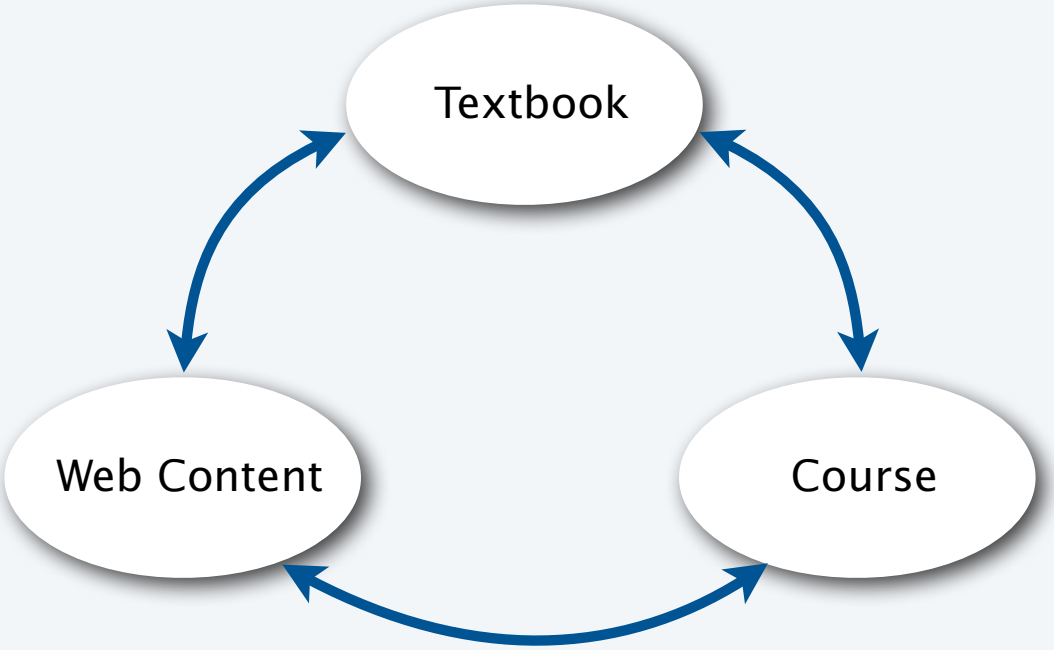
A way forward

embraces technology to integrate three abstractions that are *here to stay*:

1. A textbook for use by students to *learn and study* the details of a subject.

2. A course that encourages a community of scholars to learn together.

3. Web content for use by students to *explore and interact with* the material.



Examples (stay tuned). CS courses at Princeton implemented by RS and Kevin Wayne

The "course" abstraction

has been an essential part of education for a millenium and is *here to stay*

What is a course?

- *Lectures* to introduce to inspire.
- Assignments
- Exams
- *Precepts* to work in small groups.

U. of Bologna, founded in 1088

Purpose of a course

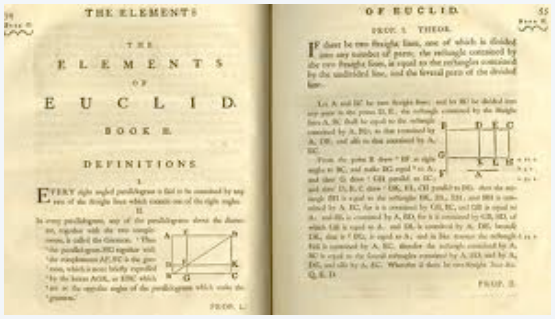
- Enable a "community of scholars" to teach and learn a subject.
- Serve as a building block in a *curriculum*.

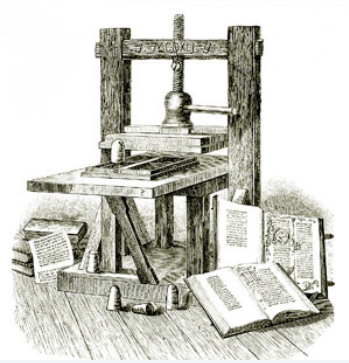
A university lecture in the 1350s

The "textbook" abstraction

has been an essential component in education for centuries and is *here to stay*

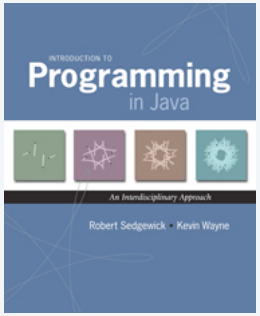
Well-understood since the Greeks.

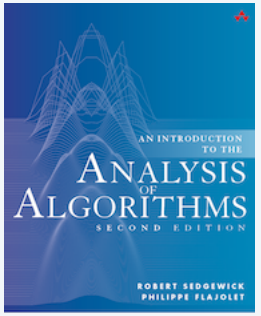


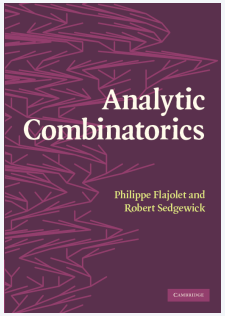


Enabled for the masses by Gutenberg.

Relevant for this talk:







Purpose of a textbook

- Articulate what students can reasonably learn about a subject in a semester.
- Provide a reference point for future studies related to the subject.

The "web content" abstraction

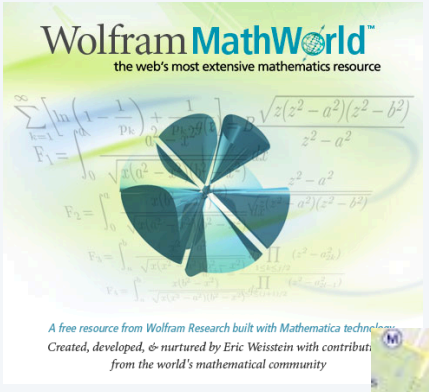
is emerging as an essential component in education and is *here to stay*

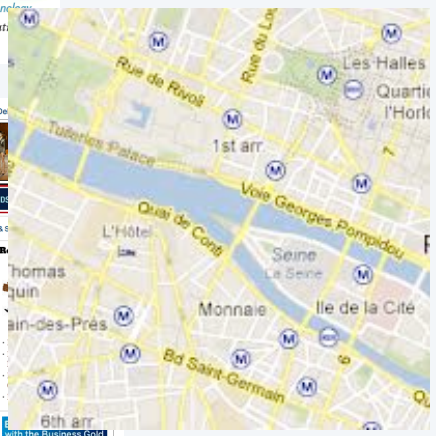
What is web content?

- *Full coverage* integrated with *web search*.
- Always up to date (*dynamic*).
- Content types *not available* in print.

Issues

- Basic properties still evolving.
- Free? Who pays?
- Who creates it? Who maintains it?



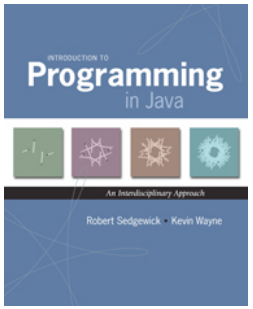


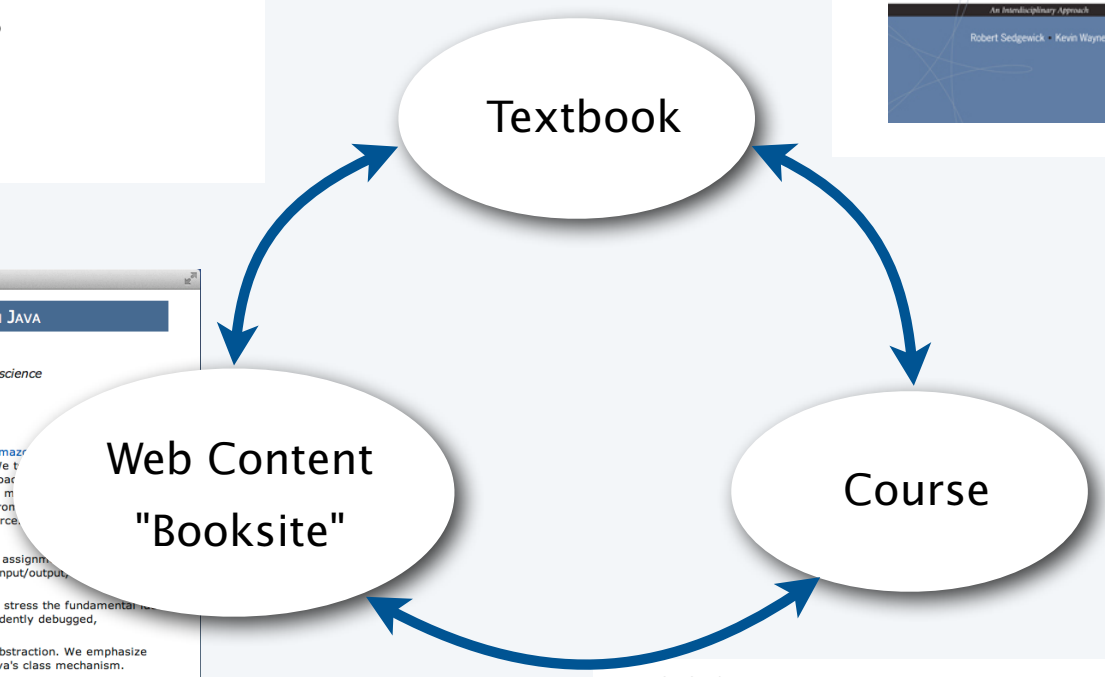
Example 1: Introduction to CS

2000s

- Raw material, digests for "story"
- Enrichment materials
- *Programming model.*

2008





10000+ files

2000+ Java programs

50+ animated demos

1.2 million unique visitors in 2011

1990s

- *Lecture presentations.*
- Assignments
- Exercises and Exams
- *Precepts.*

Example 2: Algorithms, 4th edition

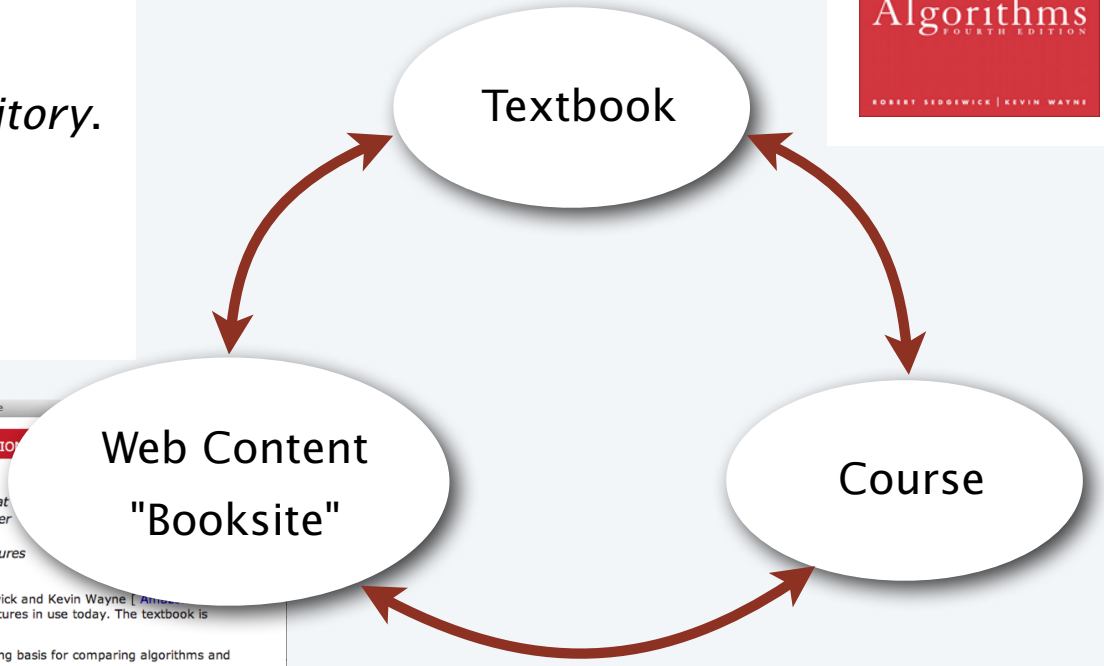
RS: Hey, maybe this could work for "Algorithms" !

KW: No problem, but...it might take some time.

2011

2011

- *Code and data repository*.
- Enrichment materials
- *Course materials*.



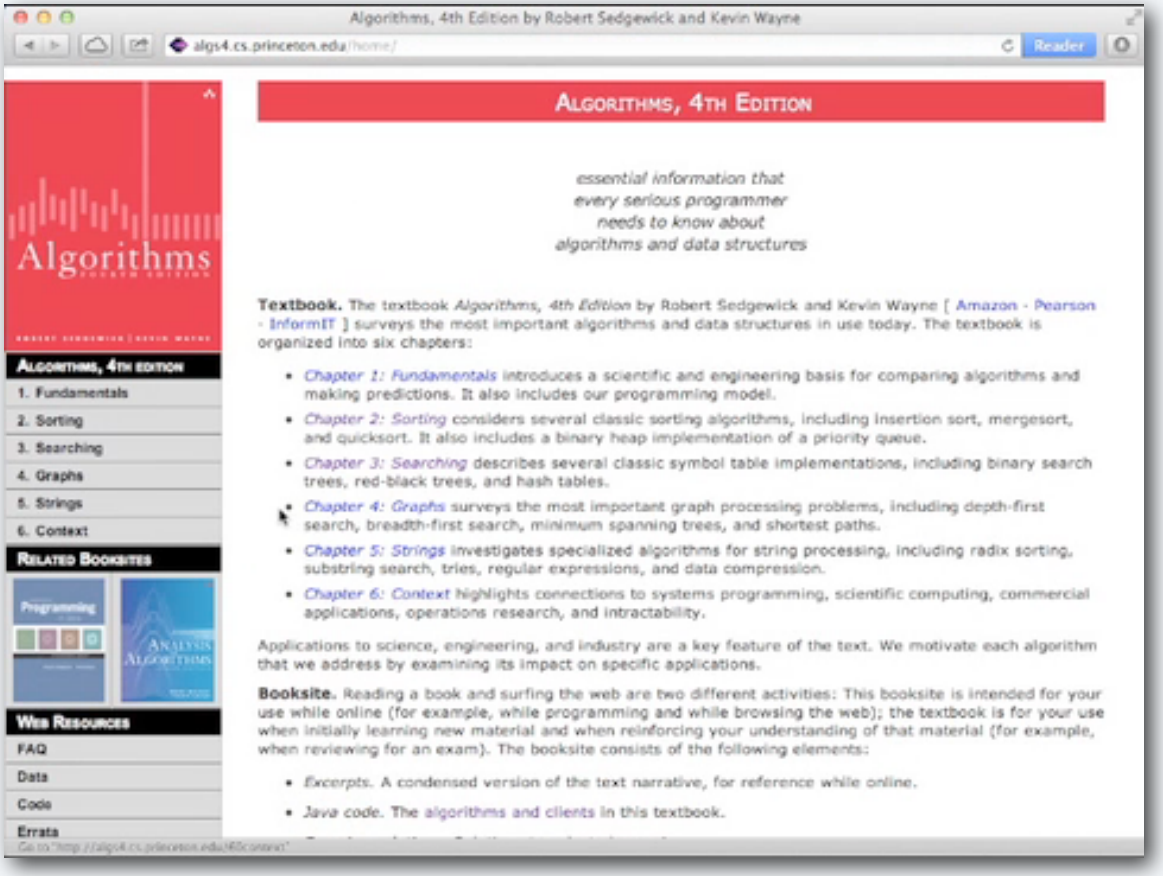
2010-2011 (4th iteration)

- *Lecture presentations*.
- Assignments
- Exercises and Exams
- *Precepts*.

Example: Algorithms booksite

Web contact associated with a book

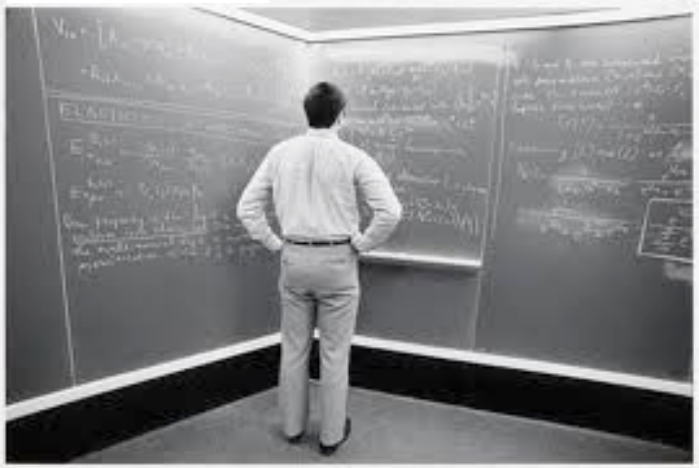
- Web presence.
- Landing and takeoff for search.
- Code, test data, animations.
- Course materials.
- A living document.
- For use while computing, exploring.



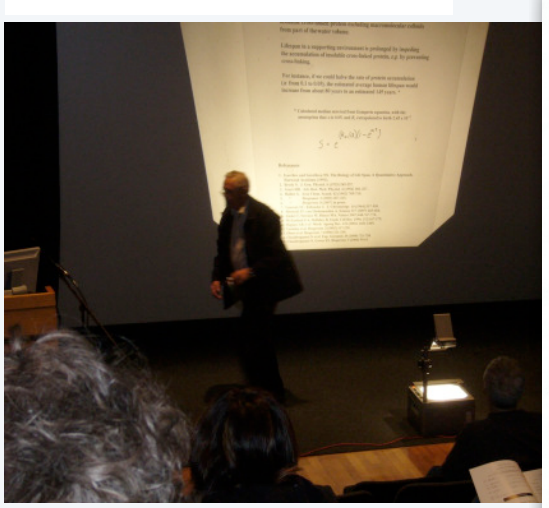
Lecture presentation materials

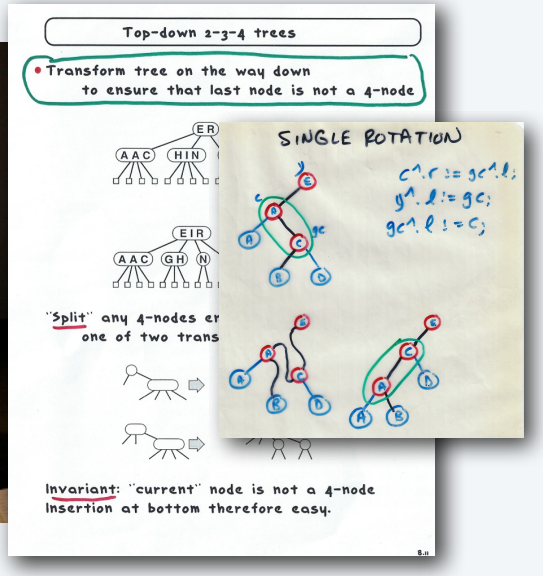
are evolving to new standard of excellence

Chalktalk



Overhead projection

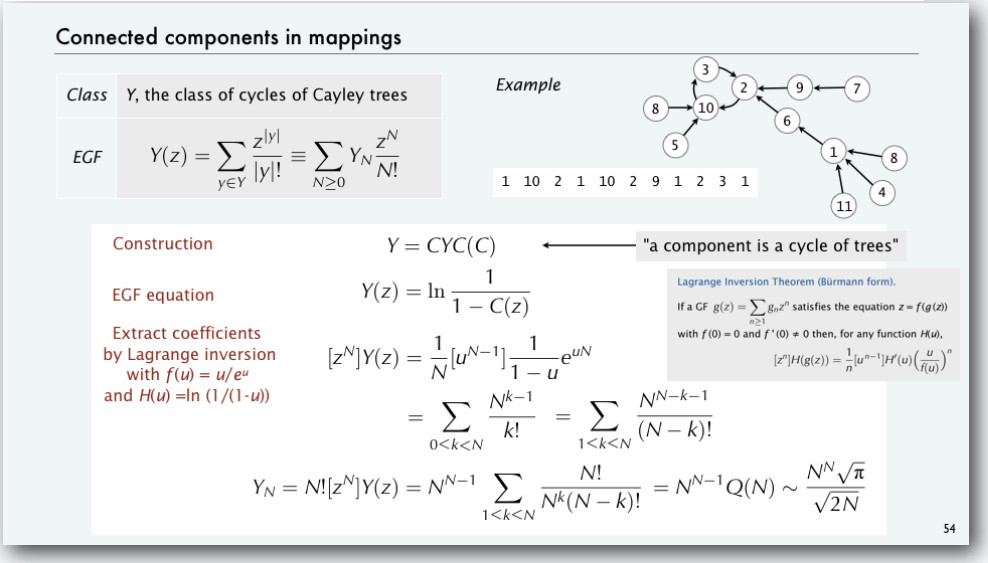




"Powerpoints"

This is a horse.

State of the art presentations (stay tuned)



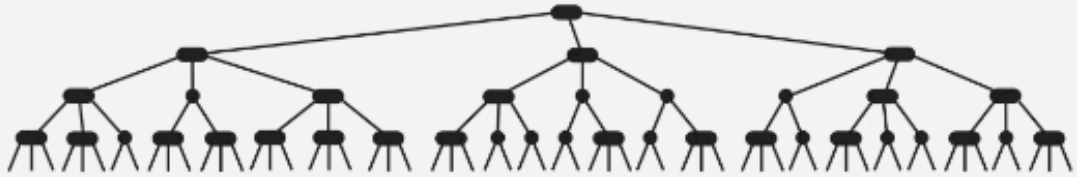
Example: An algorithms lecture

Elements

- Diagrams of data structures.
- Code.
- Animations.
- Summary Info.
- "Story".

2-3 tree: performance

Perfect balance. Every path from root to null link has same length.



Tree height.

- Worst case: $\lg N$. [all 2-nodes]
- Best case: $\log_3 N \approx .631 \lg N$. [all 3-nodes]
- Between 12 and 20 for a million nodes.
- Between 18 and 30 for a billion nodes.

Guaranteed logarithmic performance for search and insert.

13

Course materials

are dramatically improved via web dissemination

Detailed assignments

Online discussion

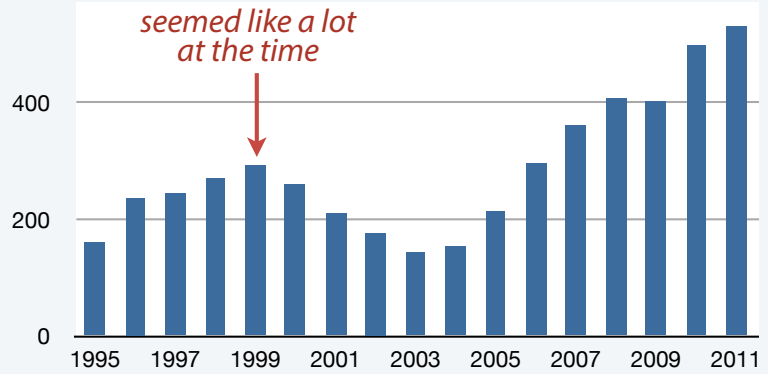
Electronic submission and assessment

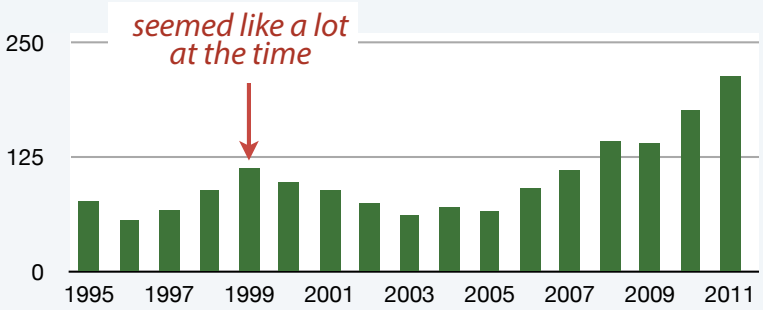
Schedule of events

2011: Time to declare victory?

Introduction to CS enrollments

- *Double* the height of the "bubble"
- 60% of all Princeton students.





"Algorithms" enrollments

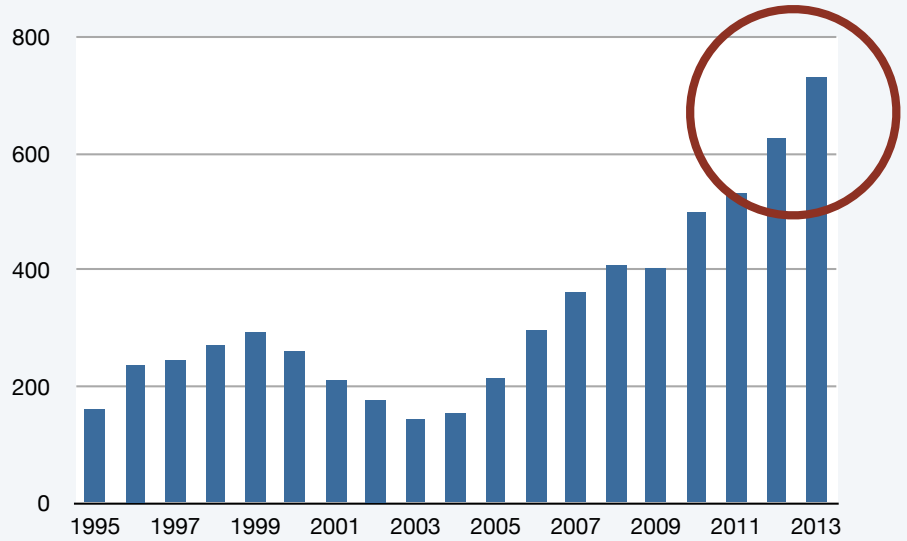
- *Three times* the height of the "bubble"
- 35% of all Princeton students.

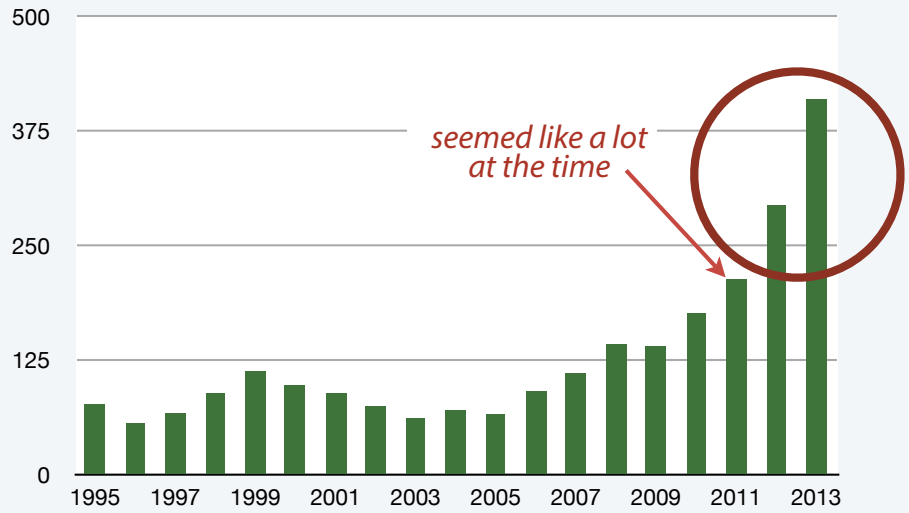
Q. (2011) When will enrollments start to decline?

2013: Time to declare victory? Not yet!

Introduction to CS enrollments

- *Triple* the height of the "bubble"
- 60% of all Princeton students.





"Algorithms" enrollments

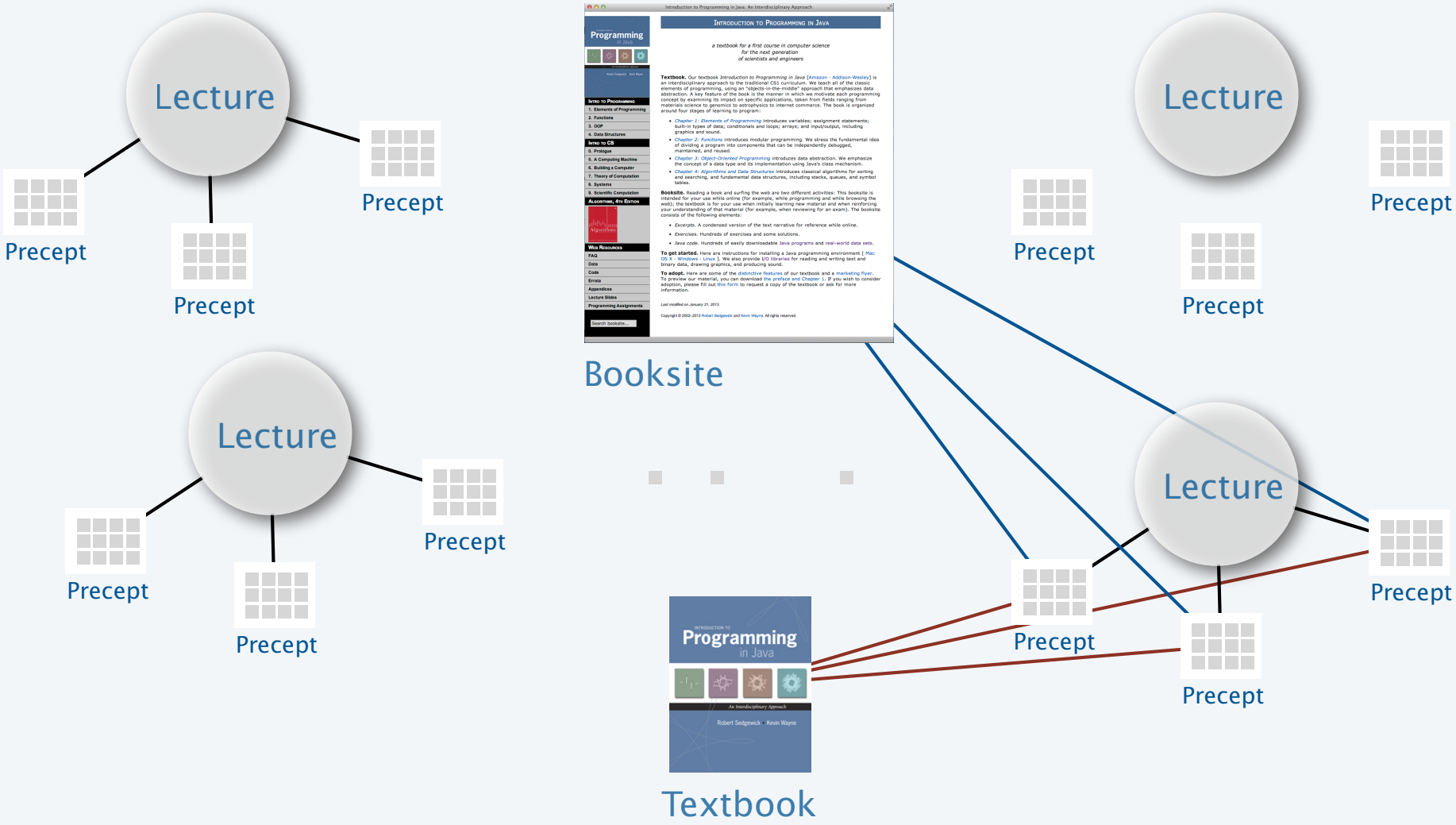
- *Four times* the height of the "bubble".
- *Doubled* in the last two years.
- 35% of all Princeton students.

Q. (2011) When will enrollments start to decline?

A. (2013) Sometime after they stop *accelerating* !

Primary advantage of booksite model

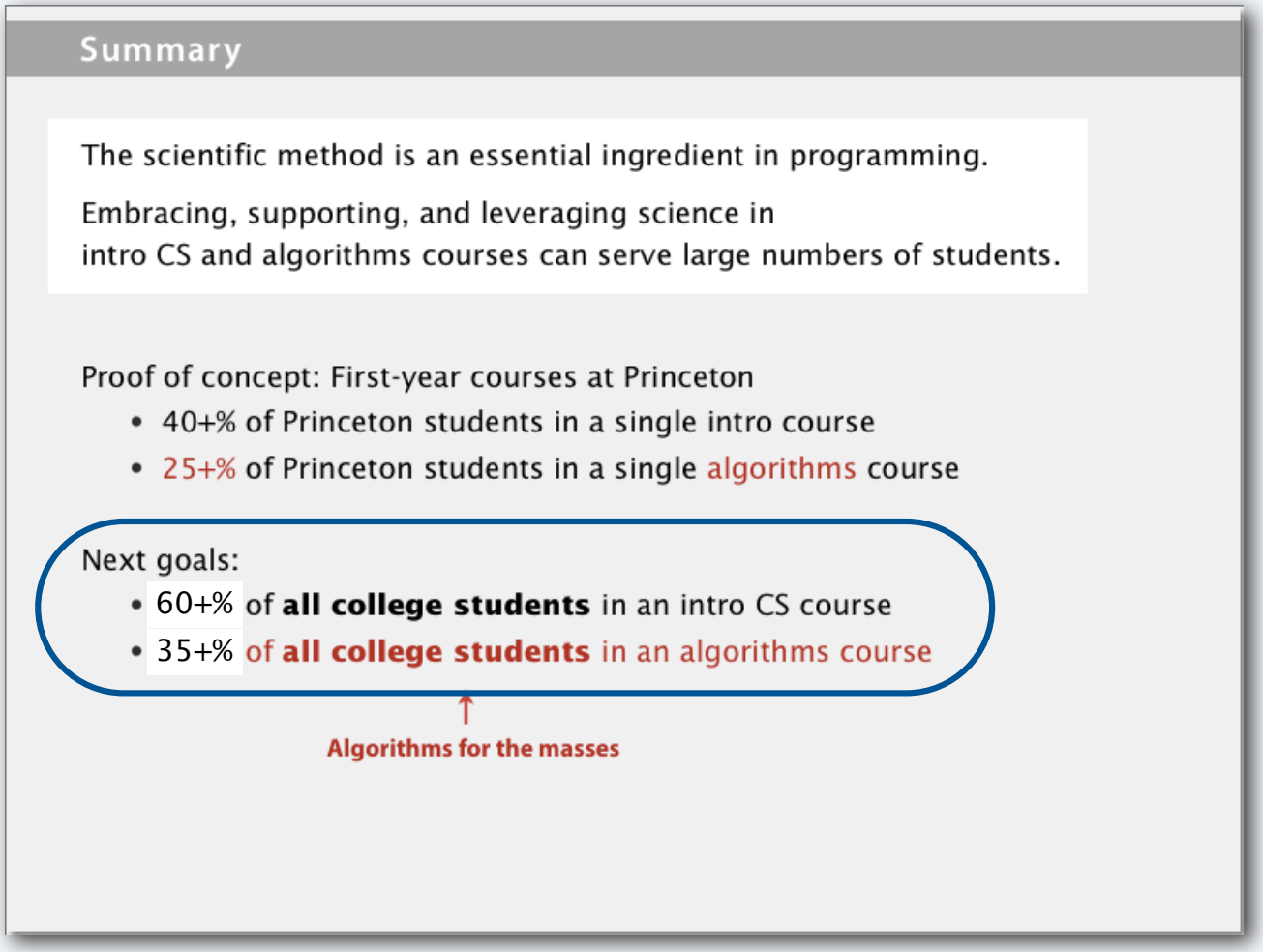
is *scalability* allowing reach to at least an order of magnitude more students



Handles huge numbers of students worldwide.
Instructors use book and booksite as basis for teaching.
Individuals are directly accessing book and booksite for self-study.

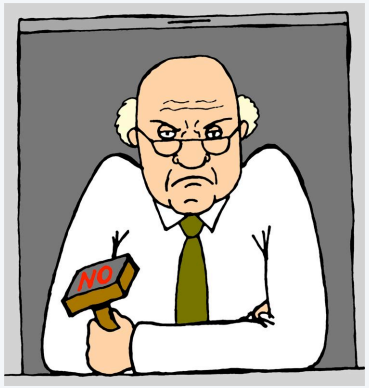
Next challenge (2011)

RS. *Algorithms for the masses* (ANALCO, San Francisco 2011)



Confession: No idea how we would get there...

2012: MOOCs go mainstream



With apologies to our actual administrators

Q. Are you interested in teaching online?

RS. No. (Too much work to do it properly.)

Q. Trustees want it, so we're doing it anyway. Are you in?

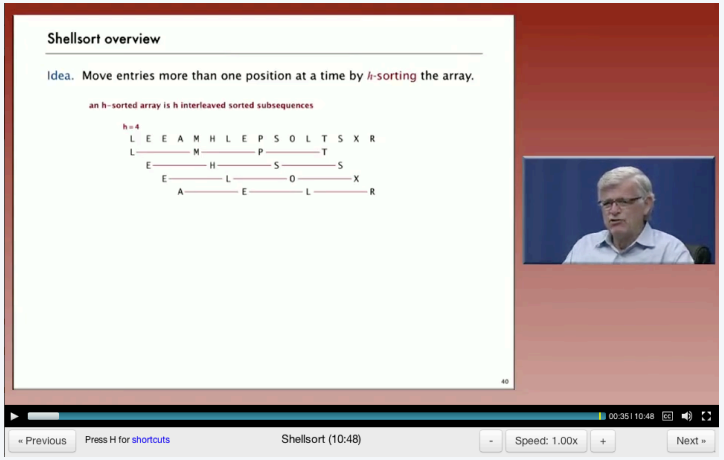
RS+KW. An offer we cannot refuse...

Andrew Ng and Daphne Koller

coursera

An online platform for the "course" abstraction

"Algorithms, Part I" (Fall 2012)



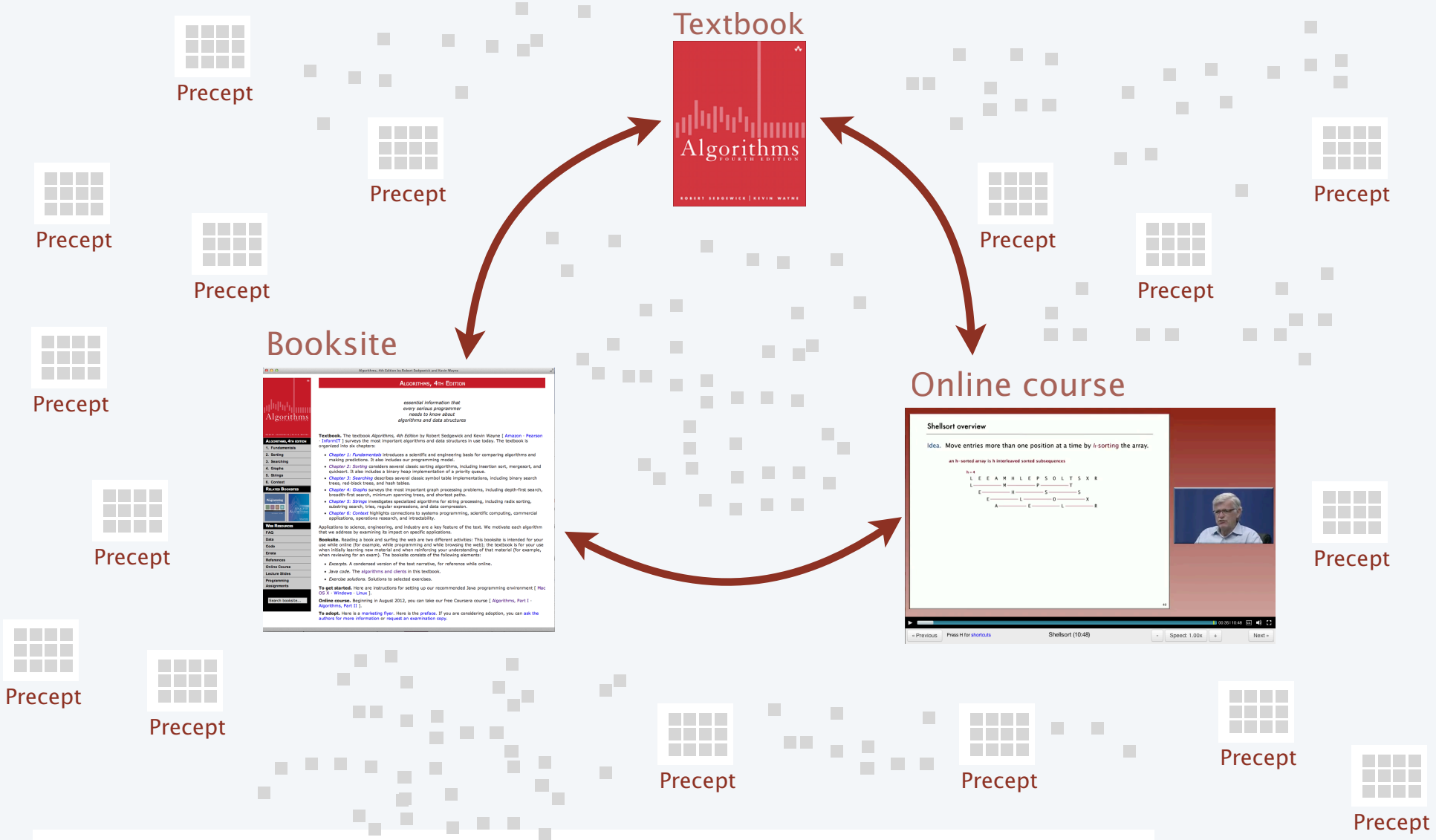
Immediate realization:

Our model is *perfectly suited* to go online.

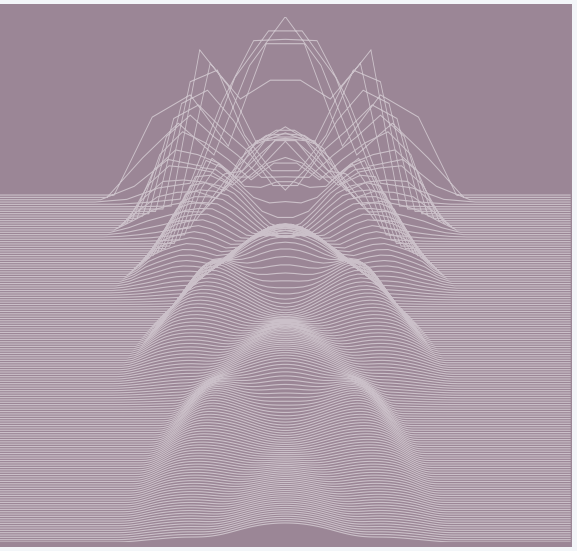
Details are another story...

Online teaching

extends our reach to at least *another* order of magnitude more students



Handles huge numbers of students worldwide.
Instructors use book, *course*, and booksite as basis for teaching.
Individuals are directly accessing all three for self-study.



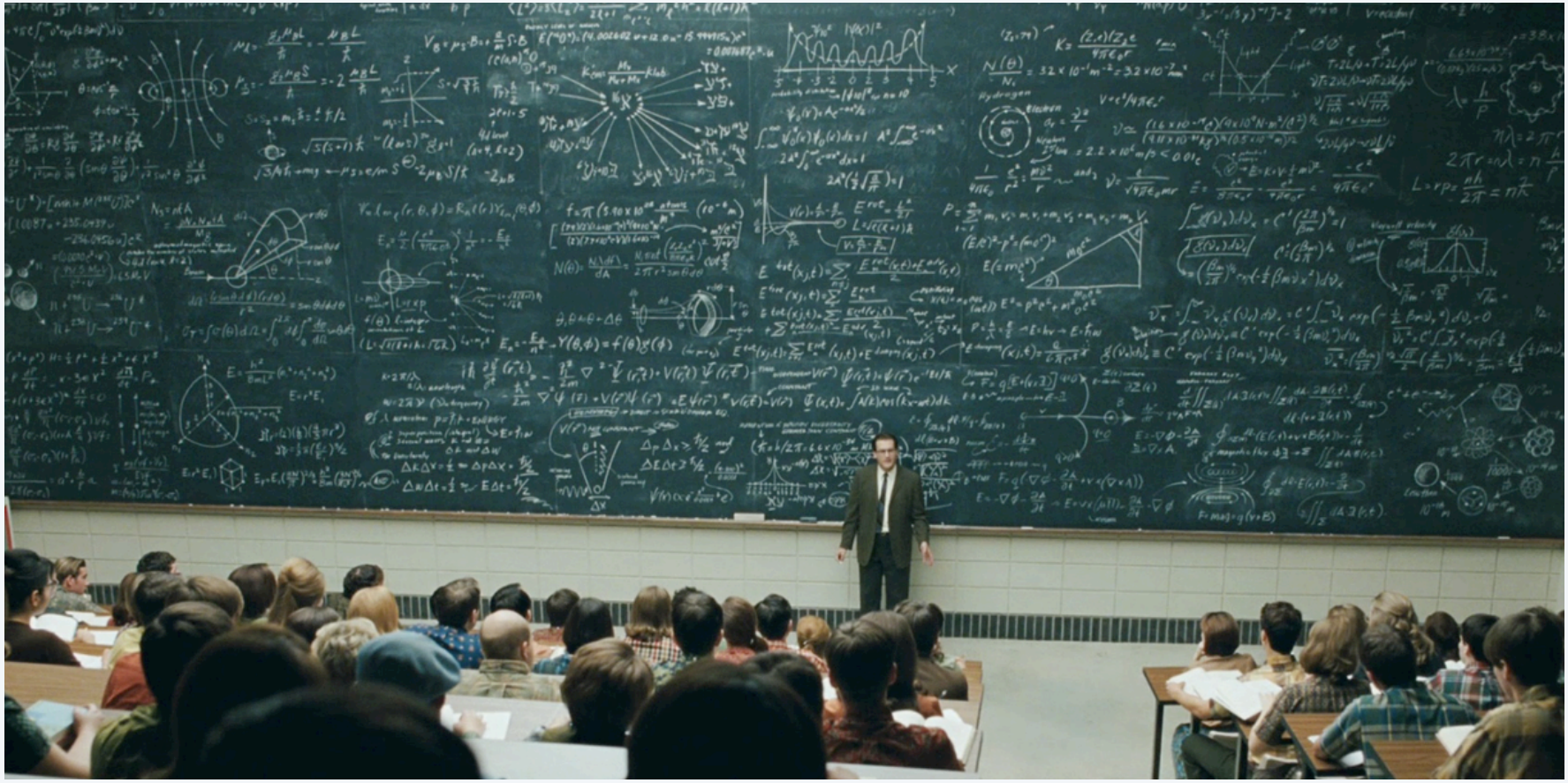
Analytic Combinatorics for the masses

- Frequently asked questions
- Disruptive changes
- A way forward
- **Analytic combinatorics**
- Questions answered

Analytic combinatorics

RS: Hey, maybe this could work for "Analytic Combinatorics" !

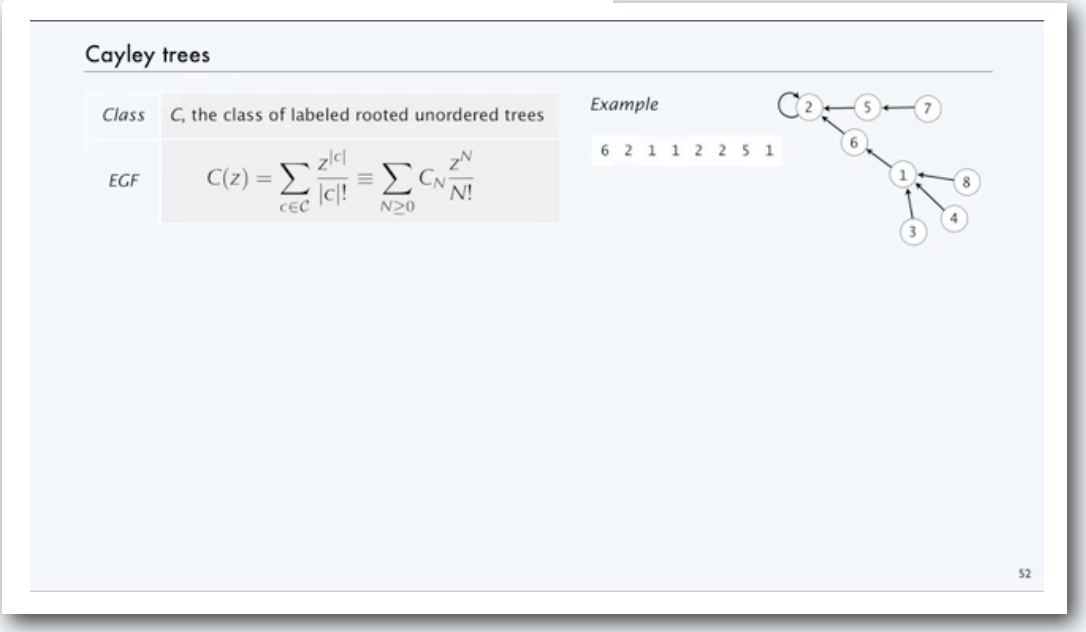
Prevailing wisdom: No way—*interactive chalktalk* is time-honored and needed for math.



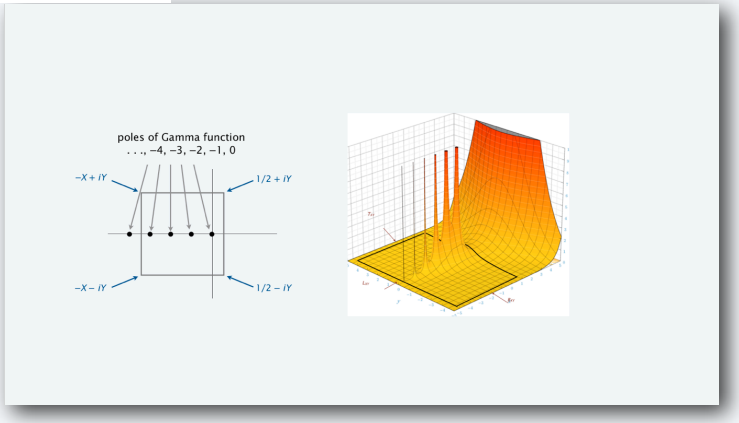
RS: Seriously?

Presentation elements: Analytic Combinatorics

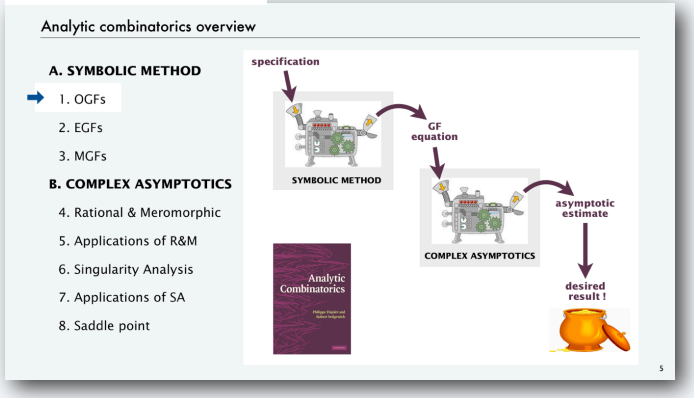
Mathematical derivations



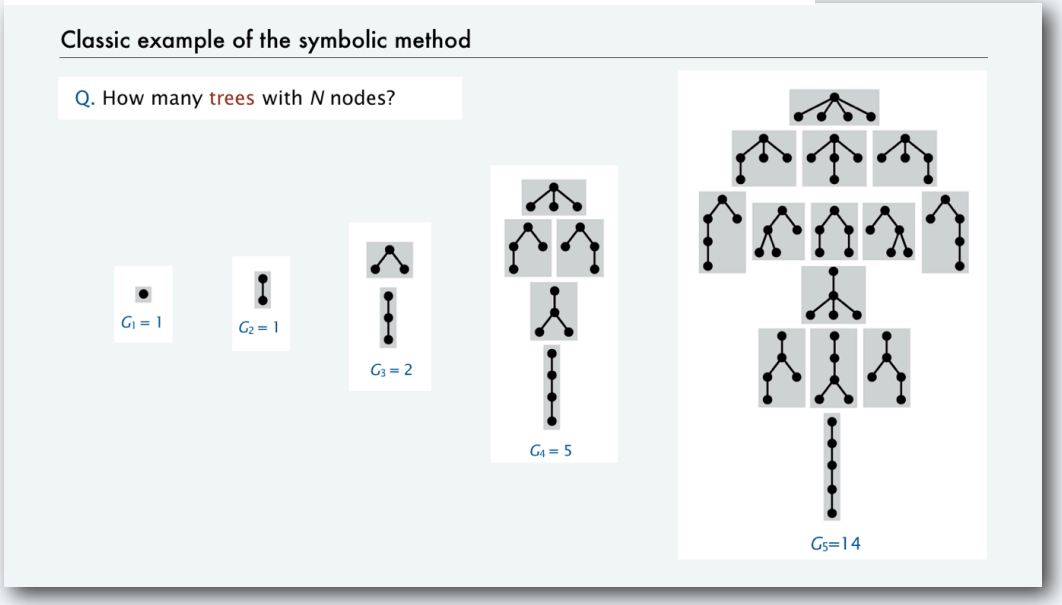
Plots



"Story"



Drawings of combinatorial objects

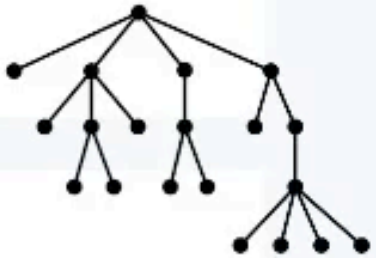


Example: An Analytic Combinatorics Presentation

Analytic combinatorics: How many trees with N nodes?

Symbolic method

Combinatorial class	G, the class of all trees
Construction	$G = \bullet \times SEQ(G)$ ← "a tree is a node and a sequence of trees"
OGF equation	$G(z) = z(1 + G(z) + G(z)^2 + G(z)^3 + \ldots) = \frac{z}{1 - G(z)}$
	$G(z) - G(z)^2 = z$
Quadratic equation	$G(z) = \frac{1 + \sqrt{1 - 4z}}{2}$



Classic next steps

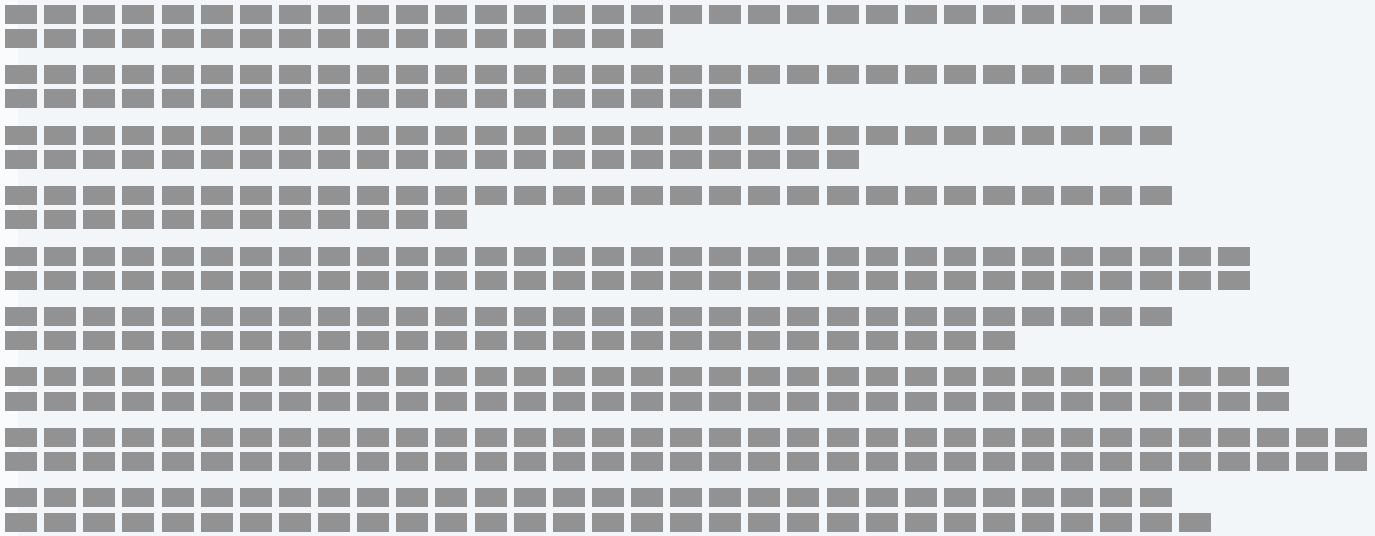
Binomial theorem	$G(z) = -\frac{1}{2} \sum_{N \geq 1} \binom{\frac{1}{2}}{N} (-4z)^N$
Extract coefficients	$G_N = -\frac{1}{2}\binom{\frac{1}{2}}{N}(-4)^N = \frac{1}{N}\binom{2N-2}{N-1} = \frac{1}{4N-2}\binom{2N}{N}$ ← detailed calculations omitted
Stirling's approximation	$\sim \frac{1}{4N} \exp\left(2N\ln(2N) - 2N + \ln\sqrt{4\pi N} - 2(N\ln(N) - N + \ln\sqrt{2\pi N})\right)$
Simplify	$G_N \sim \frac{4^{N-1}}{\sqrt{\pi N^3}}$

14

"An Introduction to Analytic Combinatorics" Lectures

Part I: Analysis of Algorithms

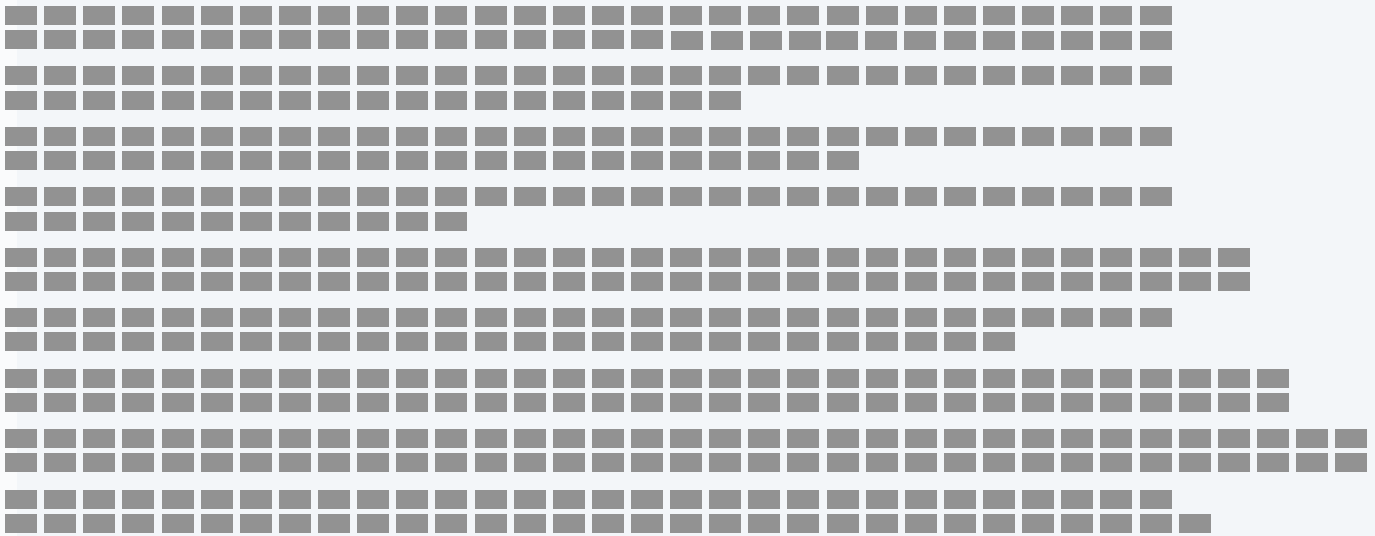
1. Introduction
2. Recurrences
3. Generating Functions
4. Asymptotic Analysis
5. Anaytic combinatorics
6. Trees
7. Permutations
8. Strings and Tries
9. Words and Mappings



~500 *slides*

Part II: Analytic Combinatorics (in development)

1. Ordinary GFs
2. Exponential GFs
3. Bivariate GFs
4. Meromorphic Asymptotics
5. MA applications
6. Singularity Analysis
7. SA Applications
8. Saddle Point
9. Advanced Topics



~500 *slides*

2013: Mission accomplished

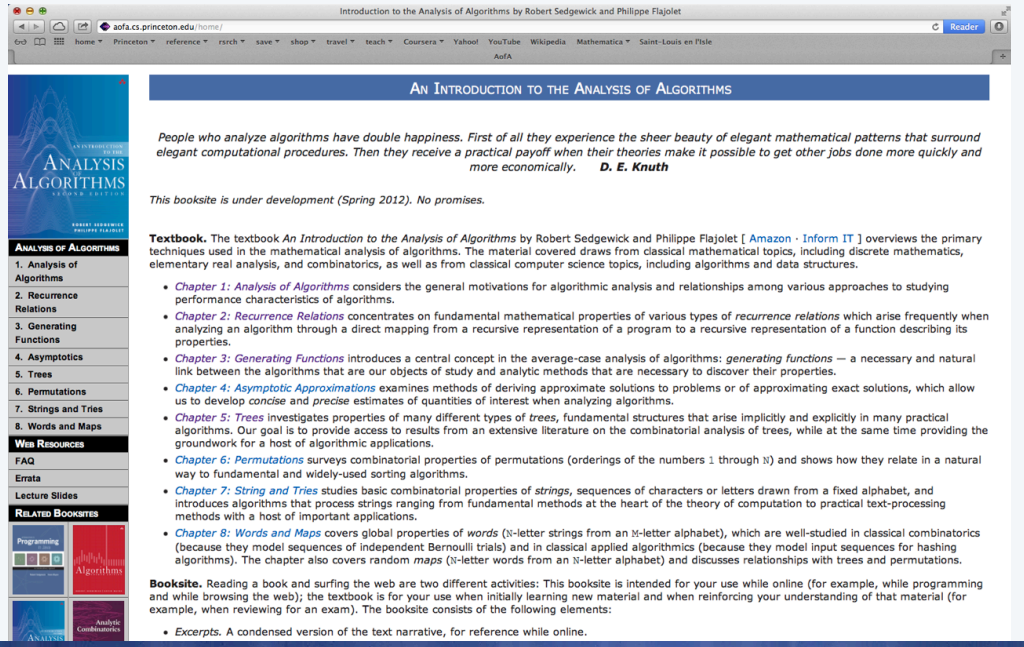
web content

`aofa.cs.princeton.edu`

text digests
code
exercise solutions
lecture slides

`ac.cs.princeton.edu`

to launch spring 2013



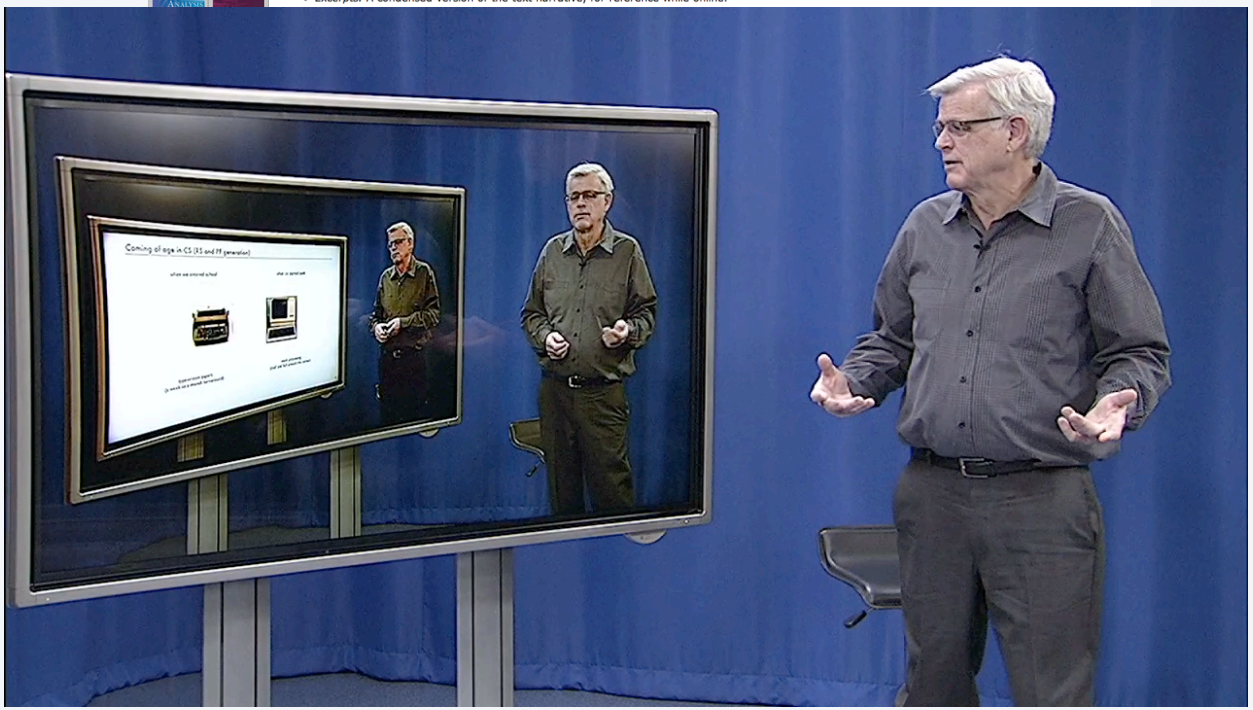
online course

Analytic Combinatorics

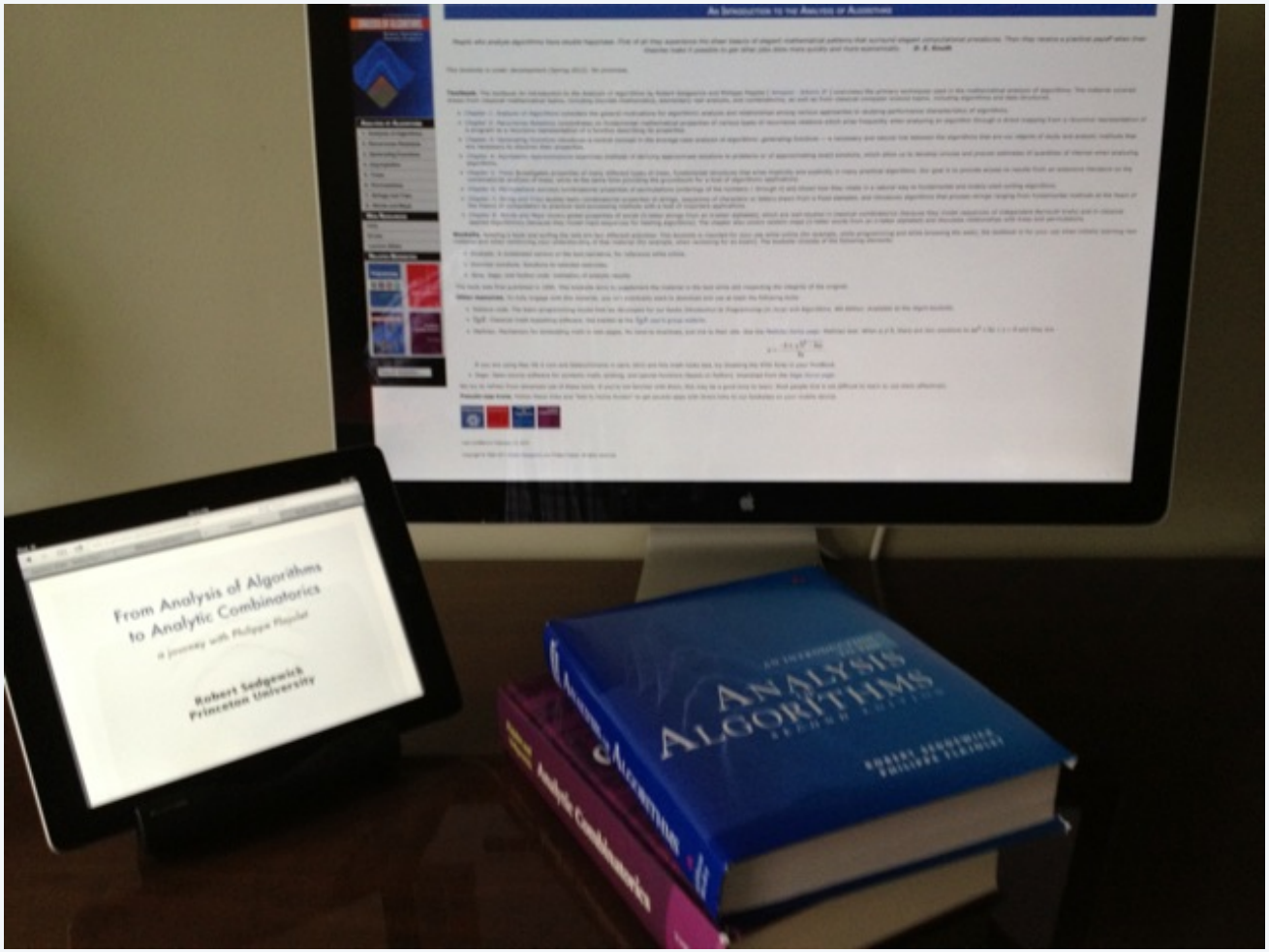
February-April 2013

10 lectures on AofA
10 lectures on AC

25,000+ registrants



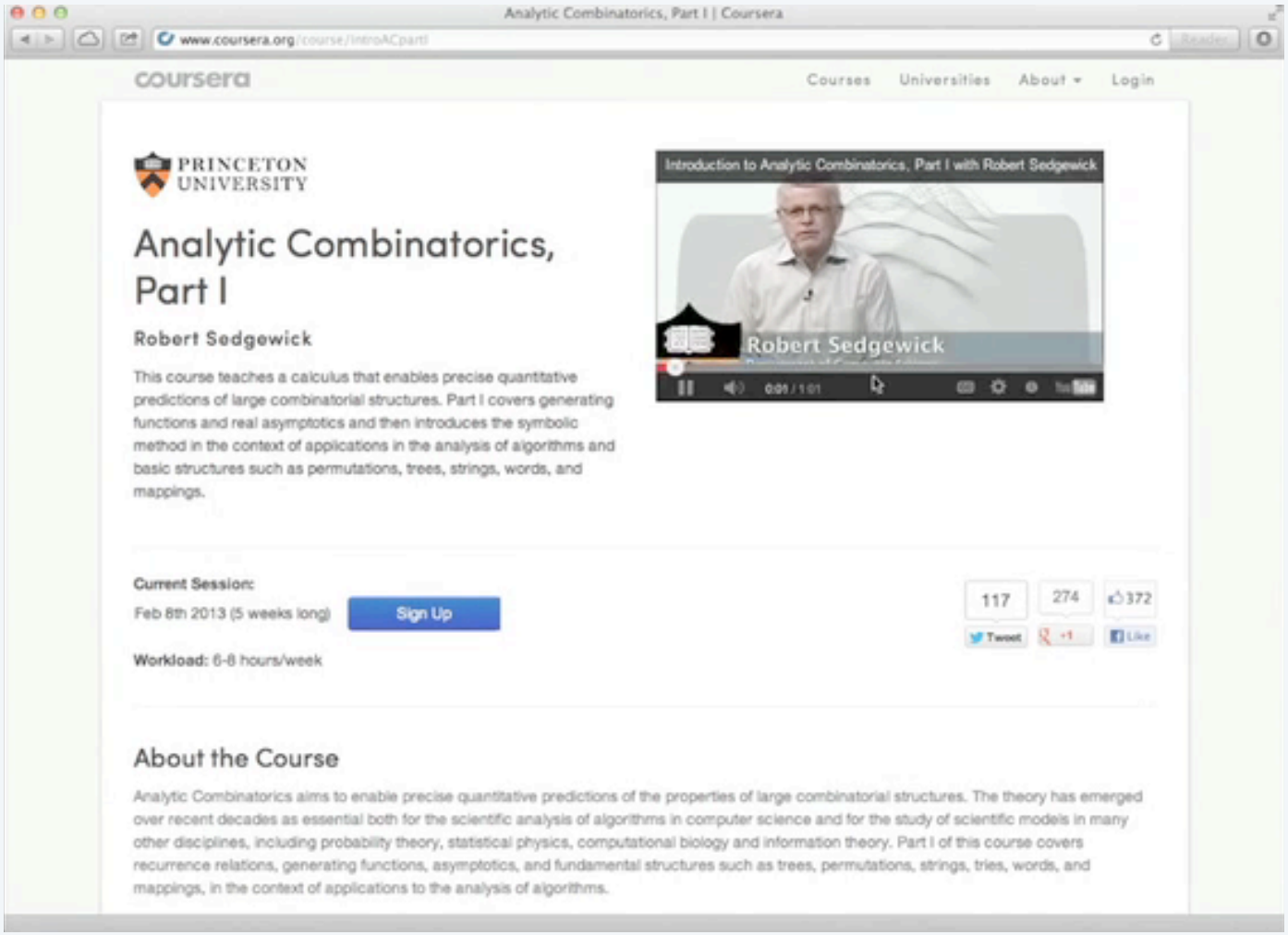
2013: Mission accomplished!

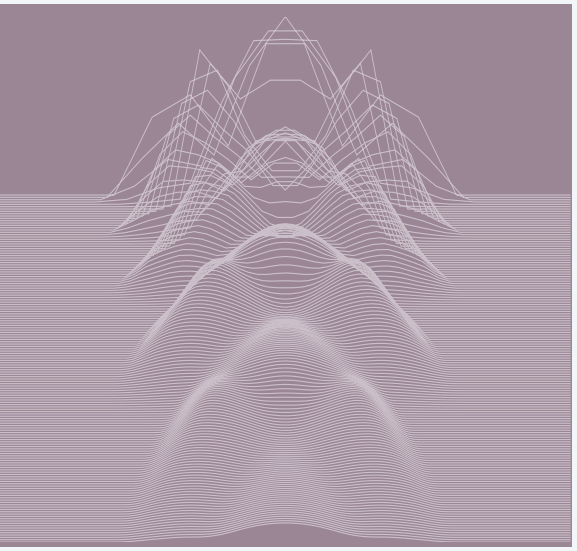


Q. Can an advanced subject such as *Analytic Combinatorics* be taught effectively online?

A. Absolutely! Indeed, advanced subjects may be a *sweet spot* for MOOCs.

How do I take an online course?





Analytic Combinatorics for the masses

- Frequently asked questions
- Disruptive changes
- A way forward
- Analytic combinatorics
- Questions answered

Questions answered

Q. What is a MOOC?

A. An *irresistible opportunity* for individuals and institutions who want to disseminate knowlege to *vastly expand* their reach.

Q. Can an advanced subject such as *Analytic Combinatorics* be taught effectively online?

A. Absolutely! Indeed, such subjects may be a *sweet spot* for MOOCs.

Q. Why?

A. *Anyone* interested in disseminating knowledge can put a good course online nowadays.

A. No university can afford to teach *all* advanced subjects.

Questions answered

With apologies to our actual students

Q. How much does it cost for me to take a MOOC?

A. It *must* be *free*. It not, it's not a MOOC.

Q. Can I get credit for taking a MOOC?

A. For *free*? Not likely.

A. You can pay to take a test.

A. You can take a MOOC as part (or all) of a course at a university.

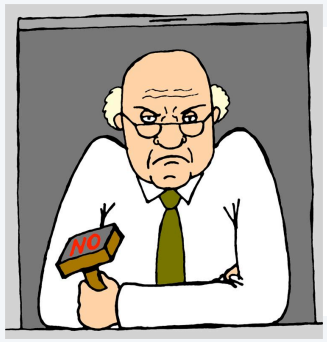
Observation: The vast majority of MOOC students don't care!

Q. If you're giving away content, what am I getting for the money at University X?

A. We make sure that you learn as much as you are able.

A. We certify that you did so.

Questions answered



Q. Can we save money by moving our teaching online?

A. NO.

A. You can *improve the quality of your teaching* and *extend your reach*.

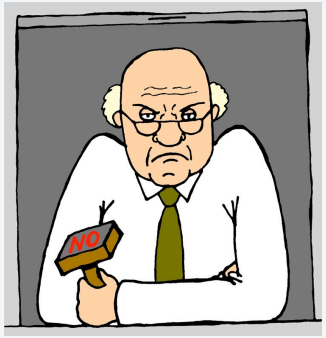
Q. How?

A. Make teachers first-class citizens.

A. Remove disincentives for researchers who choose to teach.

A. Provide support and real incentives for teaching and content creation.

Questions answered



Q. How will this impact our current students?

A. Many online courses will be better than the ones you now offer.

A. Focus will evolve to communities of scholars led by teachers working with presentations possibly developed somewhere else.

A. Everyone's education will have a significant online component.

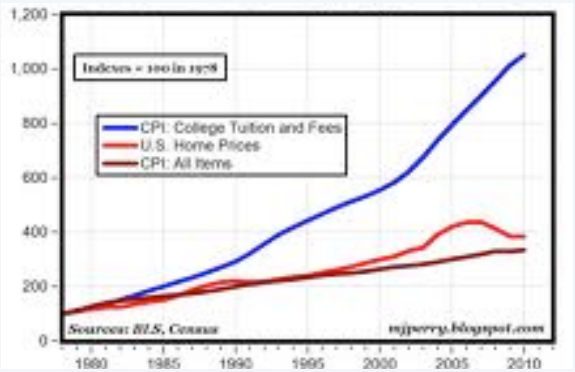
Q. How much will it cost our institution to embrace online education?

A. Less than you are spending on many things that are less central to your mission.

A. You need to plan to invest in online at the scale you are investing in the library.

A. Can you afford to *not* embrace online education?

CPI: College Tuition vs. U.S. Home Prices vs. CPI: All Items, 1978 to 2010



A parting thought

(from *John Hennessy* in an interview for an article by Ken Auletta the New Yorker, 2012)

"[Universities,] like newspapers and music companies and much of traditional media a little more than a decade ago are sailing in seemingly placid waters."

"But ... *there's a tsunami coming.*"

Analytic Combinatorics for the masses

Robert Sedgewick
Princeton University