

Big Education in the Era of Big Data

Irwin King

Associate Dean (Education), Faculty of Engineering, CUHK
Director, Shenzhen Rich Media and Big Data Analytics and Application Key Lab, SZRI, CUHK
PI, The Knowledge and Education Exchange Platform (KEEP), CUHK

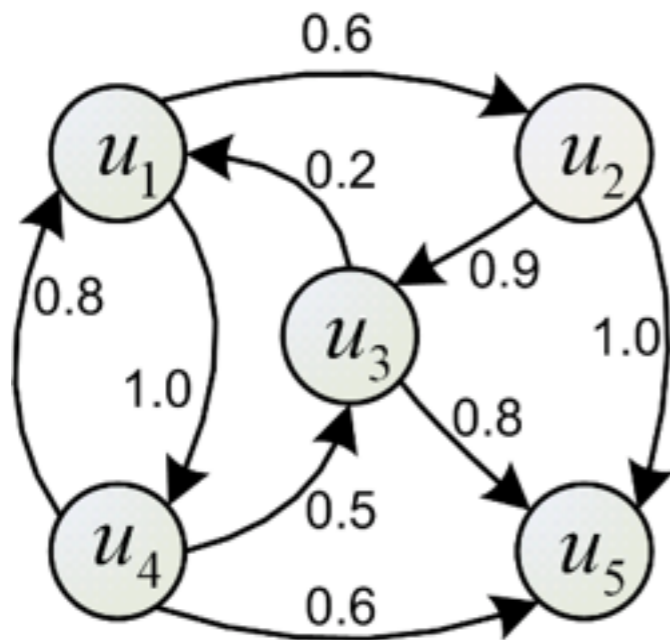
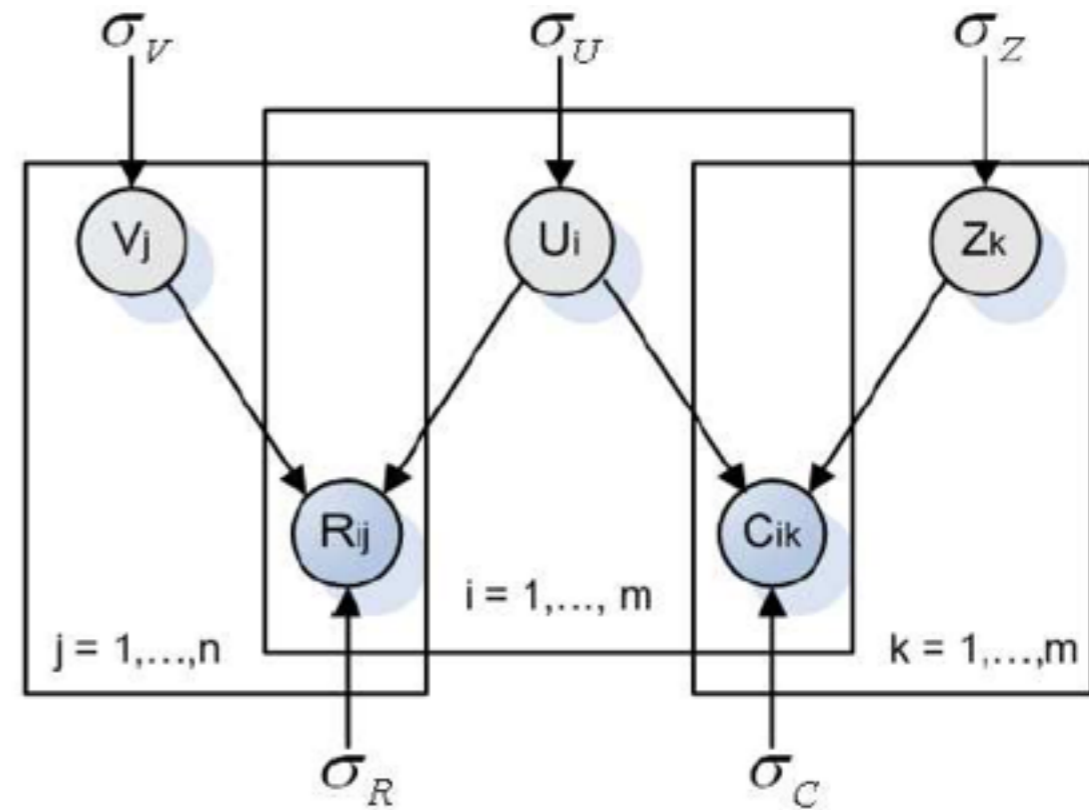
Department of Computer Science and Engineering
The Chinese University of Hong Kong

king@cse.cuhk.edu.hk
<http://www.cse.cuhk.edu.hk/~king>

©2015 Irwin King. All rights reserved.

Social Recommendations

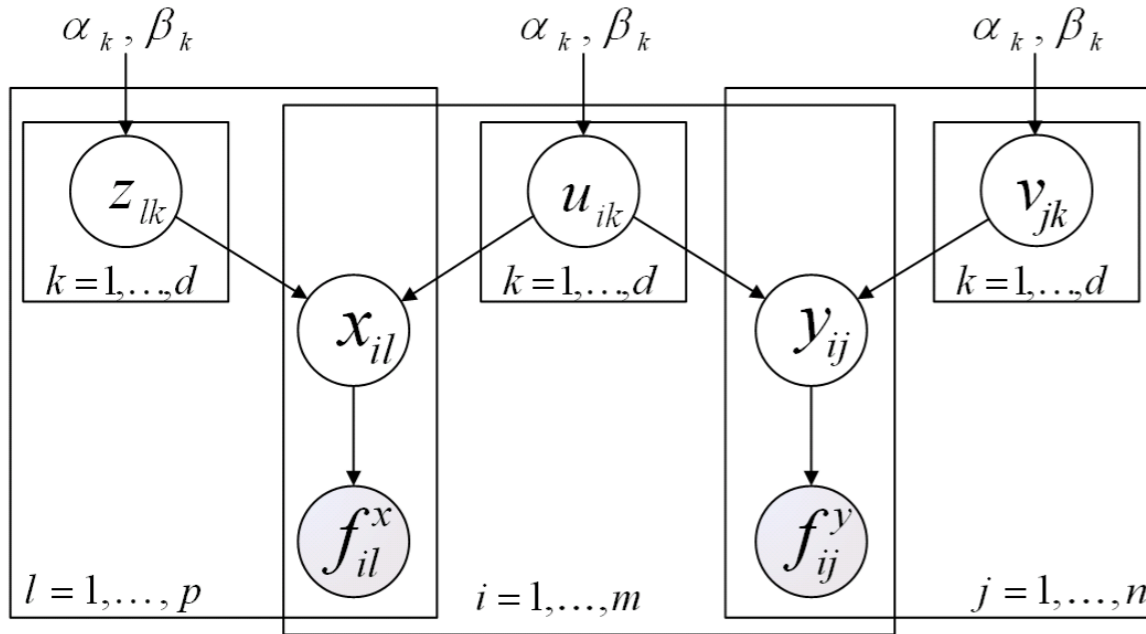
	v_1	v_2	v_3	v_4	v_5	v_6
u_1		5	2		3	
u_2	4			3		4
u_3			2			2
u_4	5			3		
u_5		5	5			3



$$\begin{aligned} \mathcal{L}(R, C, U, V, Z) = & \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (r_{ij} - g(U_i^T V_j))^2 + \frac{\lambda_C}{2} \sum_{i=1}^m \sum_{k=1}^m I_{ik}^C (c_{ik}^* - g(U_i^T Z_k))^2 \\ & + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_V}{2} \|V\|_F^2 + \frac{\lambda_Z}{2} \|Z\|_F^2, \end{aligned}$$



Collective Probabilistic Factor Model



$$\mathcal{L}(U, V, Z; F^x, F^y)$$

$$\begin{aligned}
 &= \sum_{i=1}^m \sum_{l=1}^p (f_{il}^x \ln x_{il} - x_{il}) + \sum_{i=1}^m \sum_{j=1}^n (f_{ij}^y \ln y_{ij} - y_{ij}) \\
 &+ \sum_{i=1}^m \sum_{k=1}^d ((\alpha_k - 1) \ln(u_{ik}/\beta_k) - u_{ik}/\beta_k) \\
 &+ \sum_{j=1}^n \sum_{k=1}^d ((\alpha_k - 1) \ln(v_{jk}/\beta_k) - v_{jk}/\beta_k) \\
 &+ \sum_{l=1}^p \sum_{k=1}^d ((\alpha_k - 1) \ln(z_{lk}/\beta_k) - z_{lk}/\beta_k) + \text{const.}
 \end{aligned}$$

$$u_{ik} \leftarrow u_{ik} \frac{\sum_{j=1}^n (f_{ij}^y v_{jk}/y_{ij}) + \sum_{l=1}^p (f_{il}^x z_{lk}/x_{il}) + (\alpha_k - 1)/u_{ik}}{\sum_{j=1}^n v_{jk} + \sum_{l=1}^p z_{lk} + 1/\beta_k}$$

$$v_{jk} \leftarrow v_{jk} \frac{\sum_{i=1}^m (f_{ij}^y u_{ik}/y_{ij}) + (\alpha_k - 1)/v_{jk}}{\sum_{i=1}^m u_{ik} + 1/\beta_k},$$

$$z_{lk} \leftarrow z_{lk} \frac{\sum_{i=1}^m (f_{il}^x u_{ik}/x_{il}) + (\alpha_k - 1)/z_{lk}}{\sum_{i=1}^m u_{ik} + 1/\beta_k}.$$

$$u_{ik} \leftarrow u_{ik} \frac{\theta \sum_{j=1}^n (f_{ij}^y v_{jk}/y_{ij}) + (1-\theta) \sum_{l=1}^p (f_{il}^x z_{lk}/x_{il}) + (\alpha_k - 1)/u_{ik}}{\theta \sum_{j=1}^n v_{jk} + (1-\theta) \sum_{l=1}^p z_{lk} + 1/\beta_k}$$



The grass is greener on the other side...

Be inspired!

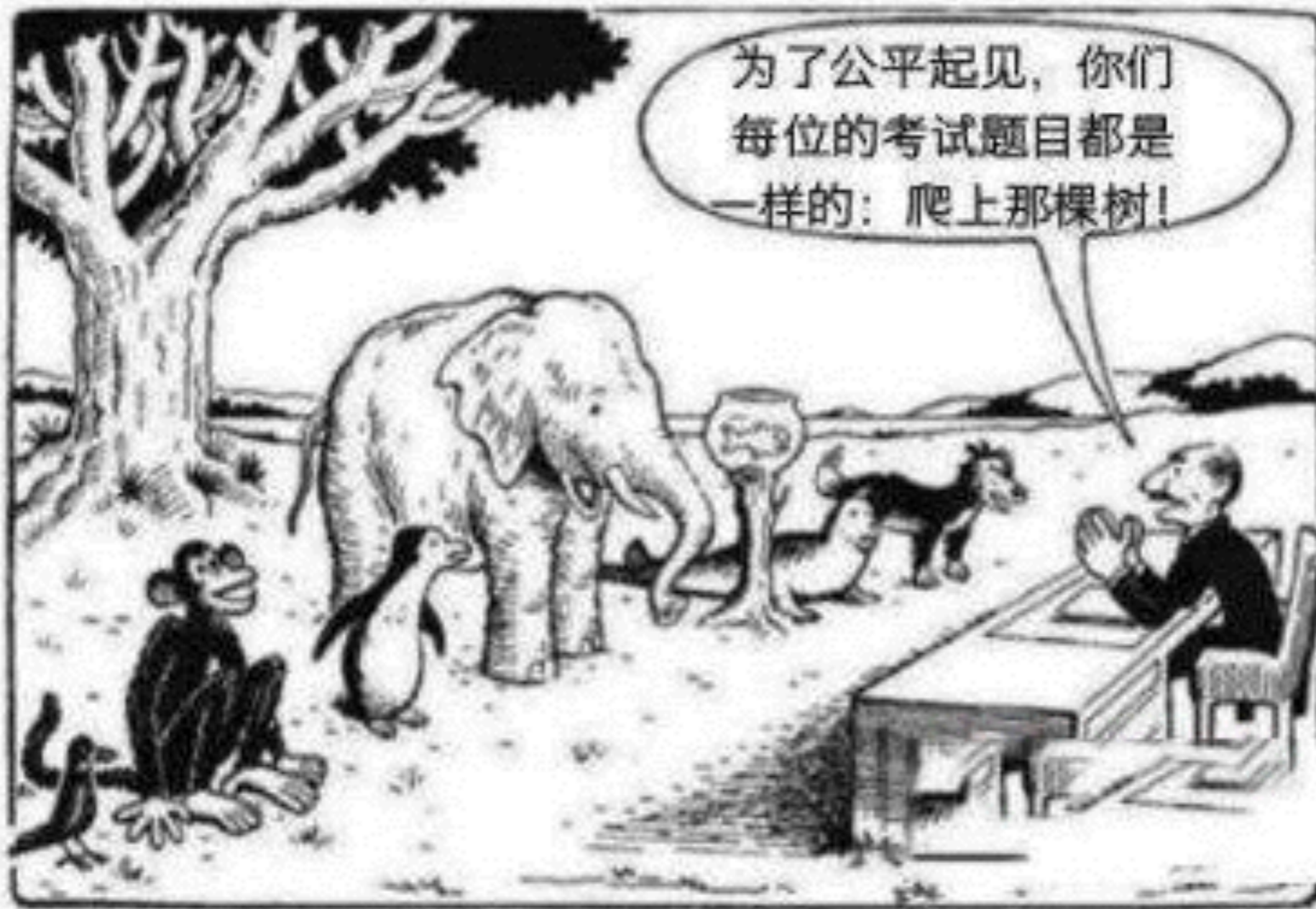
Stories and more stories...

Be informed!

The devil is in the details...

Be challenged!





我们的教育体制

GAOXIAOO.COM

Our Education System

Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.



Richard Feynman



Words of Wisdom

The **BEST** universities focus on **EDUCATION!**

The **BETTER** universities focus on
citation numbers and impact factors...

The **GOOD** universities focus on
counting the number of publications...



The First Big Data Challenge

- 1880 census
- 50 million people
- Age, gender (sex), occupation, education level, no. of insane people in household

Page No. 31
Received July 22, 1880

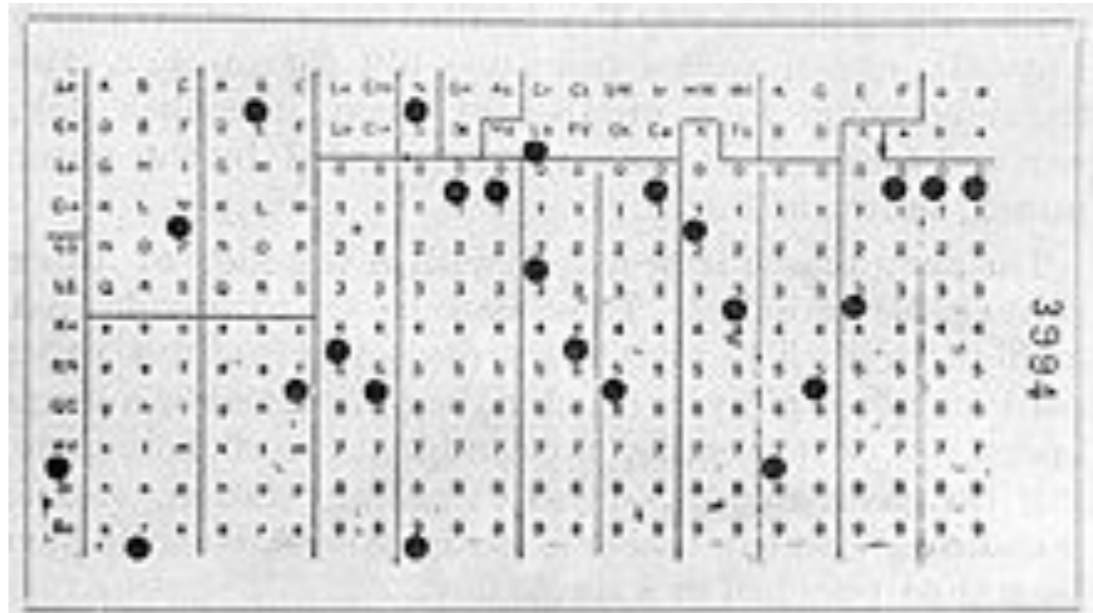
SCHEDULE 1—Inhabitants in Barton, Maryland, in the County of Washington, State of New Jersey, enumerated by me on the 1st day of June, 1880.

Name	Sex	Age	Color	Occupation
John Jones	M	25	W	Farmer
Mary Jones	F	22	W	Wife
John Smith	M	30	W	Blacksmith
Mary Smith	F	28	W	Wife
John Doe	M	40	W	Teacher
Mary Doe	F	35	W	Wife
John Brown	M	50	W	Physician
Mary Brown	F	45	W	Wife
John White	M	60	W	Blacksmith
Mary White	F	55	W	Wife



The First Big Data Solution

- Hollerith Tabulating System
- Punched cards – 80 variables
- Used for 1890 census
- 6 weeks instead of 7+ years



Big Projects

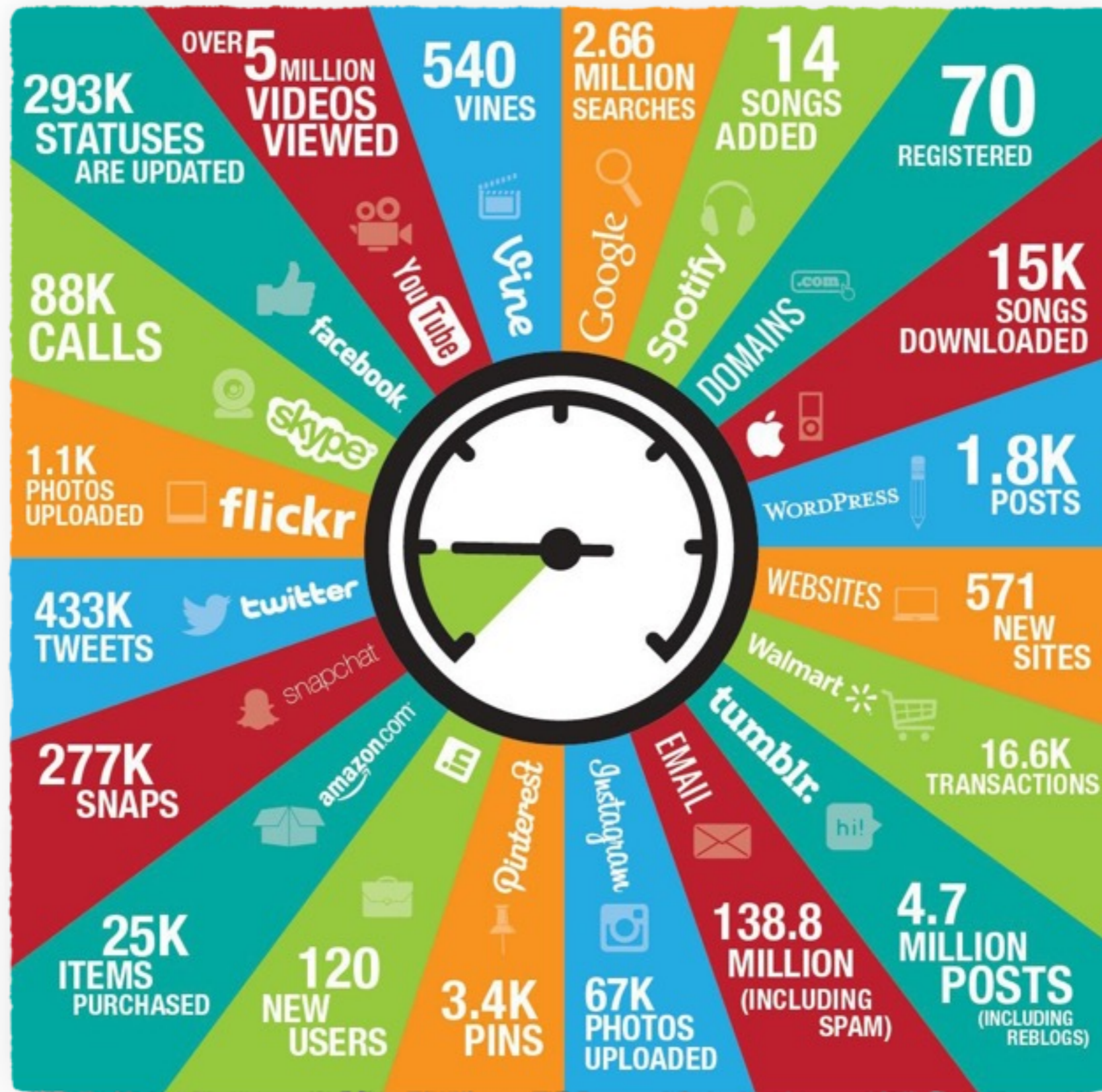
- Manhattan Project (1946 - 1949)
 - \$2 billion (approx. 26 billion in 2013)
 - Catalyst for “Big Science”
- Space Program (1960s)
 - Began in late 1950s
 - An active area of big data nowadays



Big Science

- The International Geophysical Year
 - An international scientific project
 - Last from Jul. 1, 1957 to Dec. 31, 1958
- A synoptic collection of observational data on a global scale
- Implications
 - Big budgets, Big staffs, Big machines, Big laboratories





Velocity

VALUE!

Veracity



Seven Typical Tasks

- **Querying**: spherical range-search $O(N)$, orthogonal range-search $O(N)$, nearest-neighbor $O(N)$, all-nearest-neighbors $O(N^2)$
- **Density estimation**: mixture of Gaussians, kernel density estimation $O(N^2)$, kernel conditional density estimation $O(N^3)$
- **Classification**: decision tree, nearest-neighbor classifier $O(N^2)$, kernel discriminant analysis $O(N^2)$, support vector machine $O(N^3)$, L_p SVM
- **Regression**: linear regression, kernel regression $O(N^2)$, Gaussian process regression $O(N^3)$, LASSO



Seven Typical Tasks

- **Dimension reduction:** PCA, non-negative matrix factorization, kernel PCA $O(N^3)$, maximum variance unfolding $O(N^3)$, Gaussian graphical models, discrete graphical models
- **Clustering:** k-means, mean-shift $O(N^2)$, hierarchical (FoF) clustering $O(N^3)$
- **Testing and matching:** MST $O(N^3)$, bipartite cross-matching $O(N^3)$, n -point correlation 2-sample testing $O(N^n)$, kernel embedding



Seven “Giants” of Data

- **Basic statistics:** means, covariances, etc.
- **Generalized N-body problems:** distances, geometry, etc.
- **Graph-theoretic problems:** discrete graphs
- **Linear-algebraic problems:** matrix operations
- **Optimizations:** unconstrained, convex, etc.
- **Integrations:** general dimension
- **Alignment problems:** dynamic programming, matching, etc.

Alexander Gray, GIT



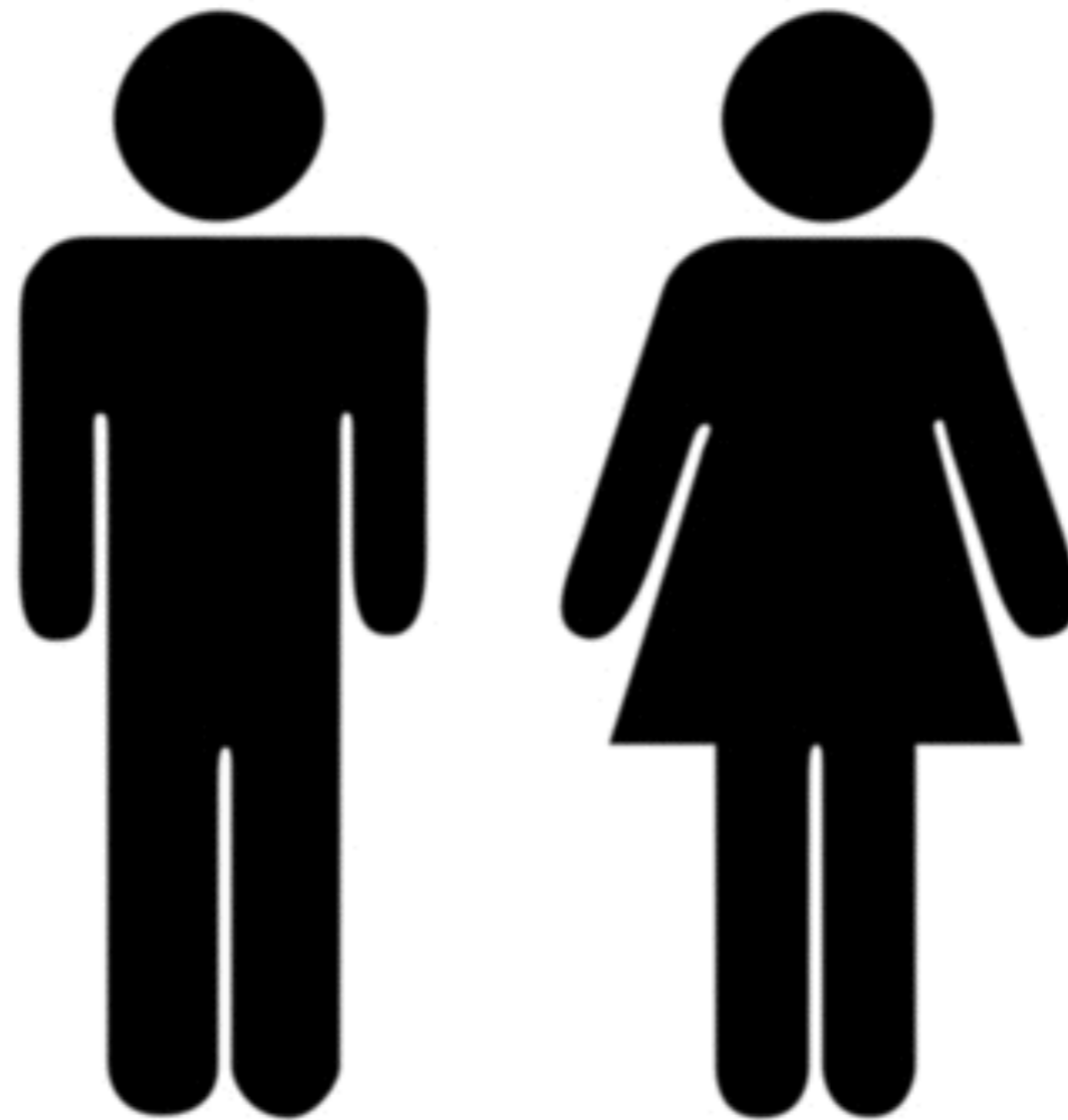
Seven General Strategies

- **Divide and conquer/indexing**: trees
- **Function transforms**: series
- **Sampling**: Monte Carlo, active learning, etc.
- **Locality**: caching, hashing, etc.
- **Streaming**: online
- **Parallelism**: clusters, GPUs, etc.
- **Problem transformation**: reformulations

Alexander Gray, GIT



What Is Big Education?



Mr. Big Data + Ms. Education



Big Education Components

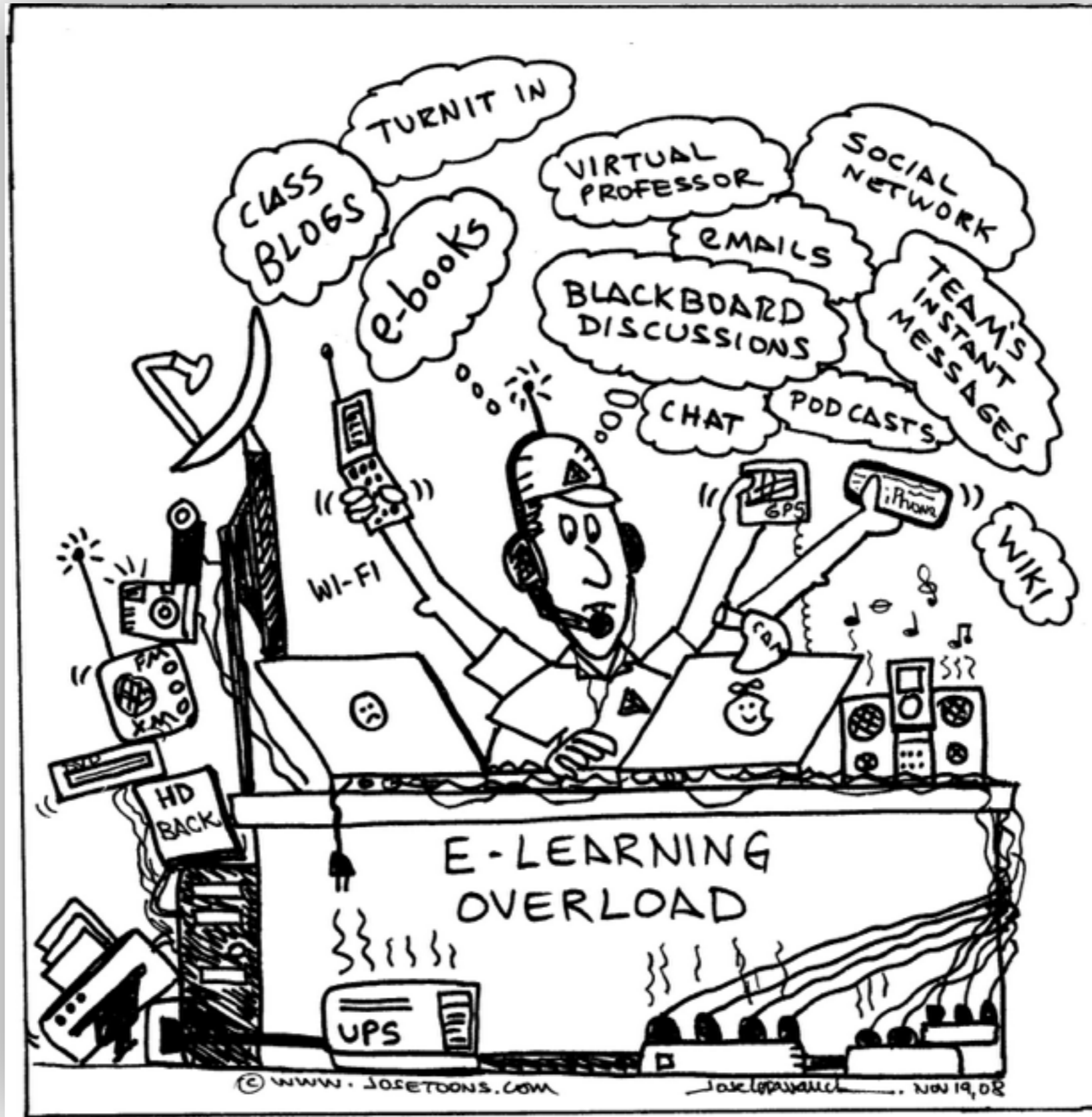
- **Pedagogy**
 - Theory and practice on education
- **Platforms and system technology**
 - Content creation, storage, delivery, etc.
 - Learning management systems
- **Algorithms for Big Education**
 - Learning analytics at scale



Big Education Applications

- **Statistics** for citation index, impact factors, h-index, bibliometrics, university ranking, etc.
- **Social computing** in students's learning networks, co-author networks, career networks, etc.
- Behavior and **learning analytics**
- Career and learning path **recommendations**
- Finding **domain experts, trendy research topics, communities**, etc.
- Identifying **plagiarism, at-risk students, frauds**, etc.
- Intelligent and automated **assessment and tutoring**







Education without values, as useful as it is, seems rather to make man a more clever devil.

C.S. Lewis



Trends in Big Education



Big Education in the Era of Big Data by Irwin King @ IES/CSBio/IAIT/iNCEB 2015, Nov. 22-25, 2015, Bangkok, Thailand





Collaboration



**Cost
Effectiveness**



Customization



The Higher Education, Continuing Education Online Learning Landscape



<http://athentica.com/wp-content/uploads/2013/10/Online-Learning-Landscape-Oct-2013.jpg>





e-Learning industry
Money spent on self-paced
eLearning across the globe

\$35.6 billion

2011

BIG MONEY

\$56.2
billion

2013

2015

it's going
to **double**
by 2015.



2013: About **4.6** out of 10 college students
are taking at least one course online.

roughly **half** of all college
classes* will be eLearning-based.



eLearning has the power to **increase information retention** rates by up to **60%** ⁽³⁾.

BIG IMPROVEMENT

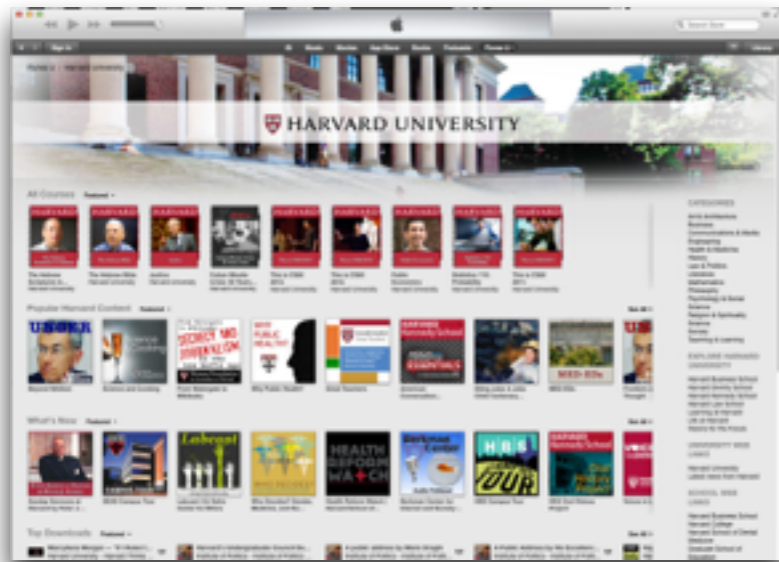
it's also **more**
effective (the learning process).



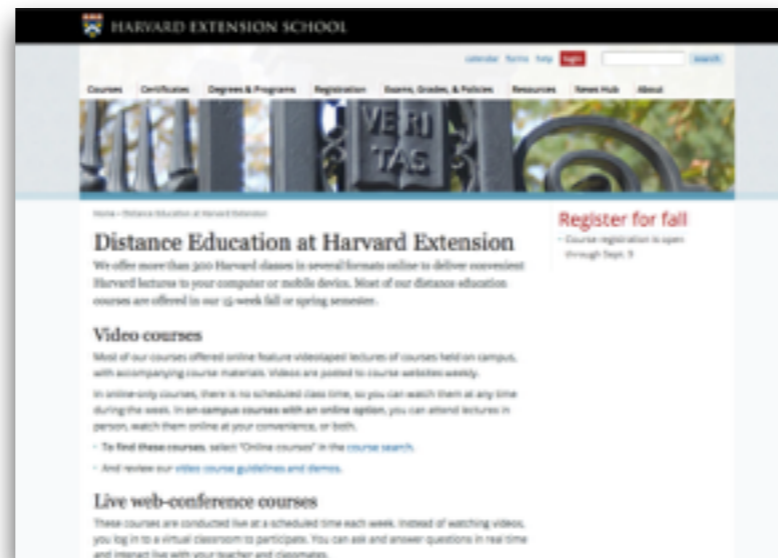
Big Education on Lifelong Learning



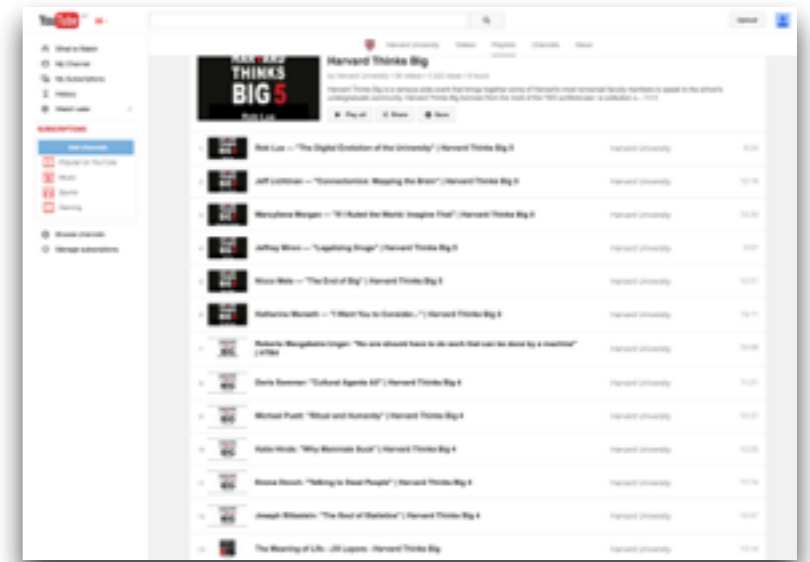
Multimodal Learning



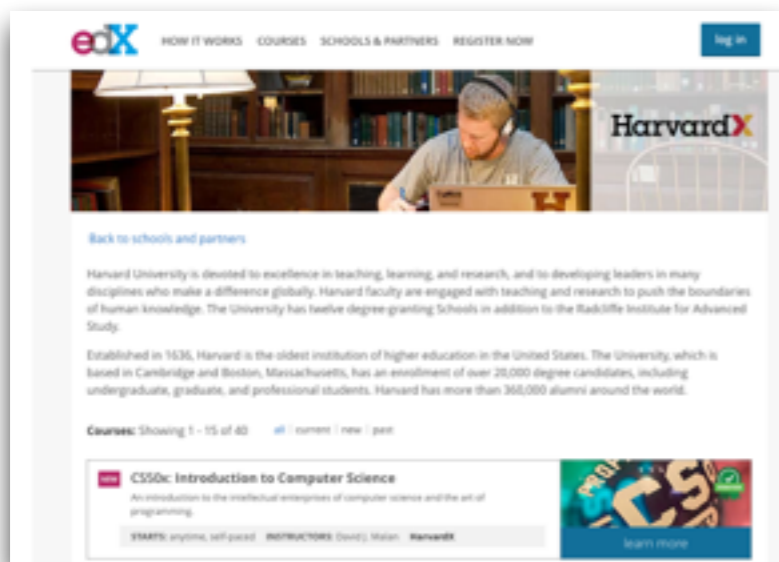
Harvard @ iTunes U



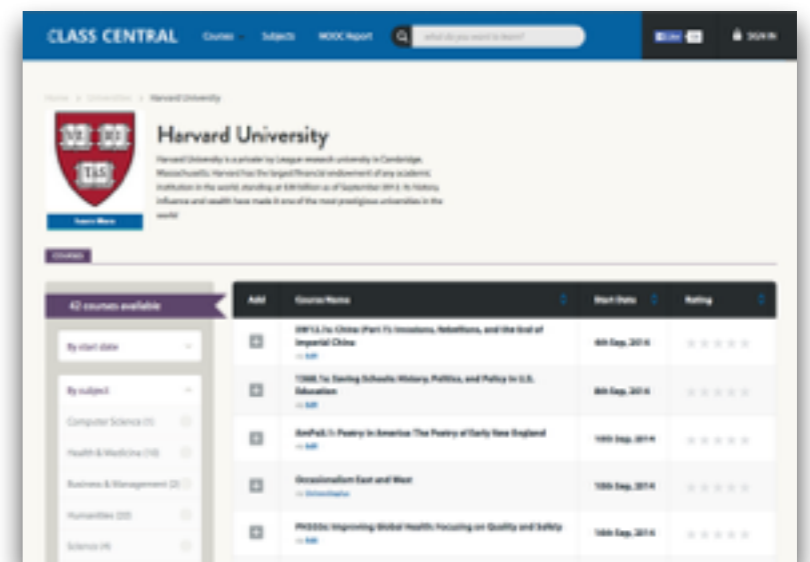
harvard.edu



Harvard @ YouTube



Harvard @ edX



Harvard @ Class Central

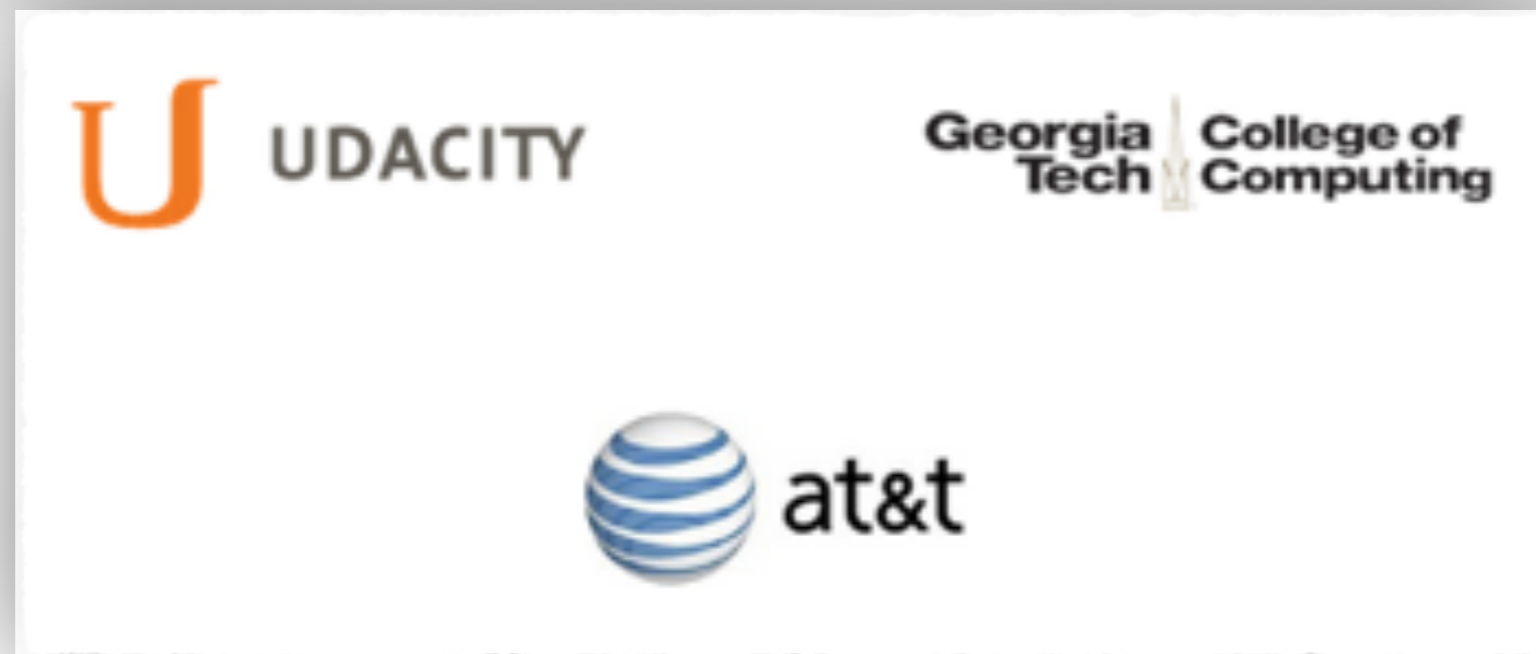


MOOC

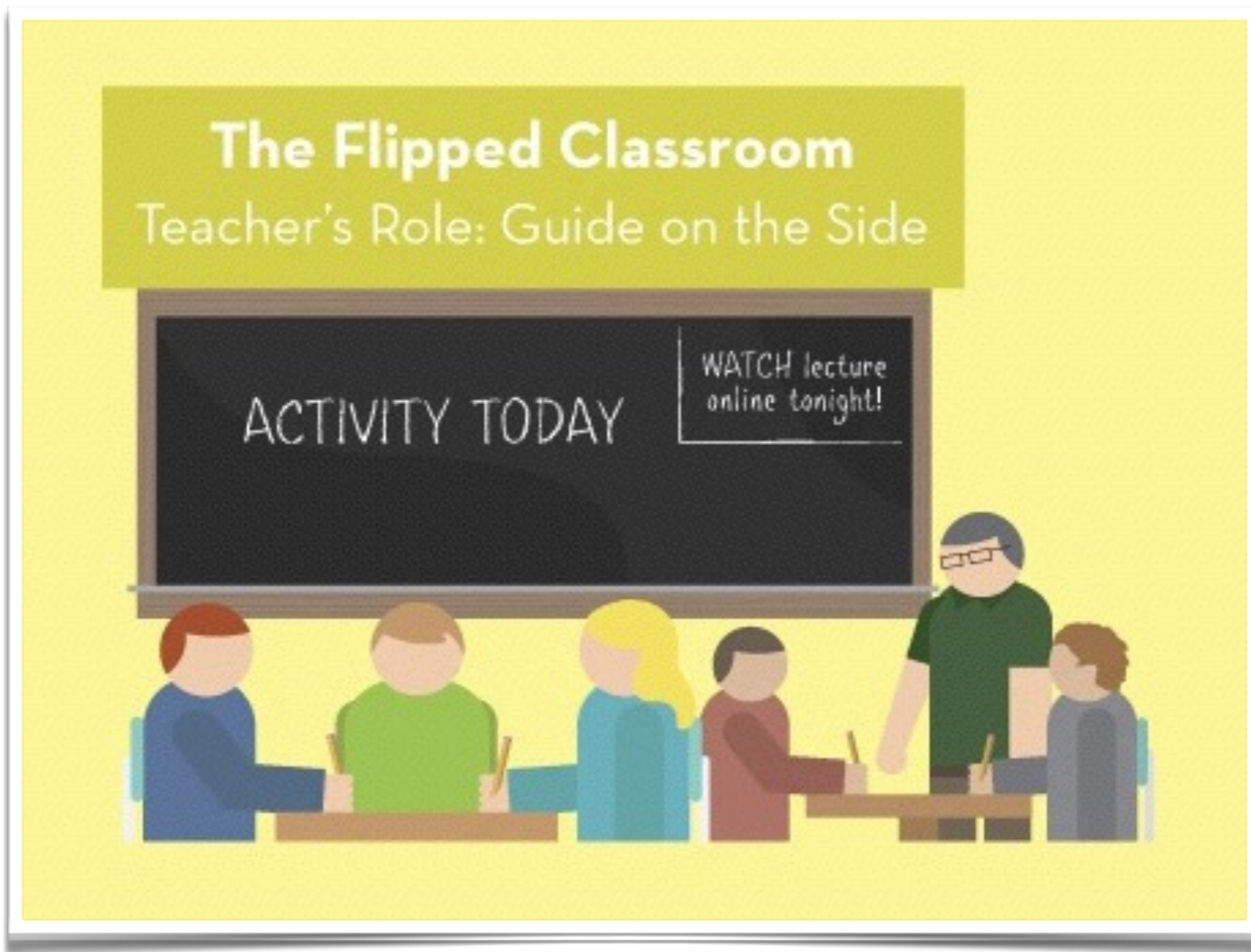
Massive Open Online Course



Small Private Online Course (SPOC) with Degree



Flipped Classroom



Computational Thinking



Barefoot would like to acknowledge the work of Julia Briggs and the eLIM team at Somerset County Council for their contribution to this poster.

Big Education in the Era of Big Data by Irwin King @ IES/CSBio/IAIT/iNCEB 2015, Nov. 22-25, 2015, Bangkok, Thailand



Microlearning

KHANACADEMY Subject: Computer pro... Coach About Donate Search for subjects, skills, and videos Log in Sign up

COMPUTER PROGRAMMING

Intro to JS: Drawing & Animation

In these tutorials, you'll learn how to use the JavaScript language and the ProcessingJS library to create fun drawings and animations. If you've never programmed before, start here to learn how!

- + Create Program
- Documentation
- ? Help Requests
- Project Evaluations
- Community Questions

ALL CONTENT IN "INTRO TO JS: DRAWING & ANIMATION"

Intro to programming

If you've never been here before, check out this introductory video first. Then get coding!

- ▶ What is Programming?
- ▶ A Tour of Programming on Khan Academy

Drawing basics

We'll show you the basics of programming and how to draw shapes.

- ▶ Intro to Drawing
- ★ Challenge: H for Hopper
- ▶ More Drawing!
- ★ Challenge: Simple Shapes!
- ★ Challenge: CRAZY Face

Coloring

We'll show you how to color and outline your shapes!

- ▶ Intro to Coloring
- ★ Challenge: Ice Cream Code
- ★ Challenge: It's a Beautiful Day
- ▶ The Power of the Docs
- ✓ Project: What's for Dinner?

Variables

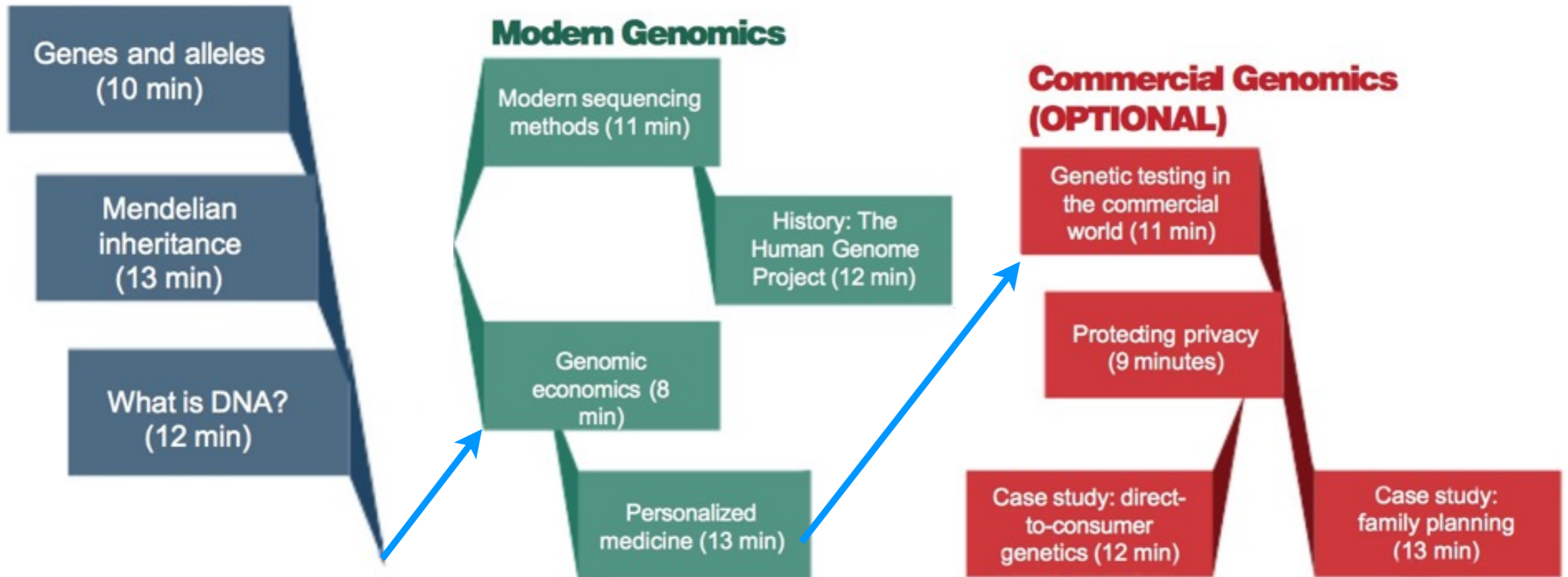
We'll cover how to use variables to hold

- ▶ Intro to Variables



Personalized Learning

Basic Genetics Refresher (OPTIONAL)



Active Learning

The screenshot shows a courseware interface with a sidebar on the left and a main content area. The sidebar contains a navigation menu with the following items:

- Week 1
- Week 2
 - Class 4: Functions, Programs, Commands
 - Class 5: Selection statements: if, switch
 - Class 6: Loops: for, nested
 - Class 7: Loops: while, vectorizing, timing code
- Week 3
- Week 4
- Week 5
- Week 6
- Worksheets

The main content area displays a video player with the title "FOR LOOP COMBINATIONS". The video content shows a slide titled "Nested loop trace example" with the following code:

```
for i = 1:3
    fprintf('*')
    for j = 1:5
        fprintf('%d', j)
    end
    fprintf('\n')
end
```

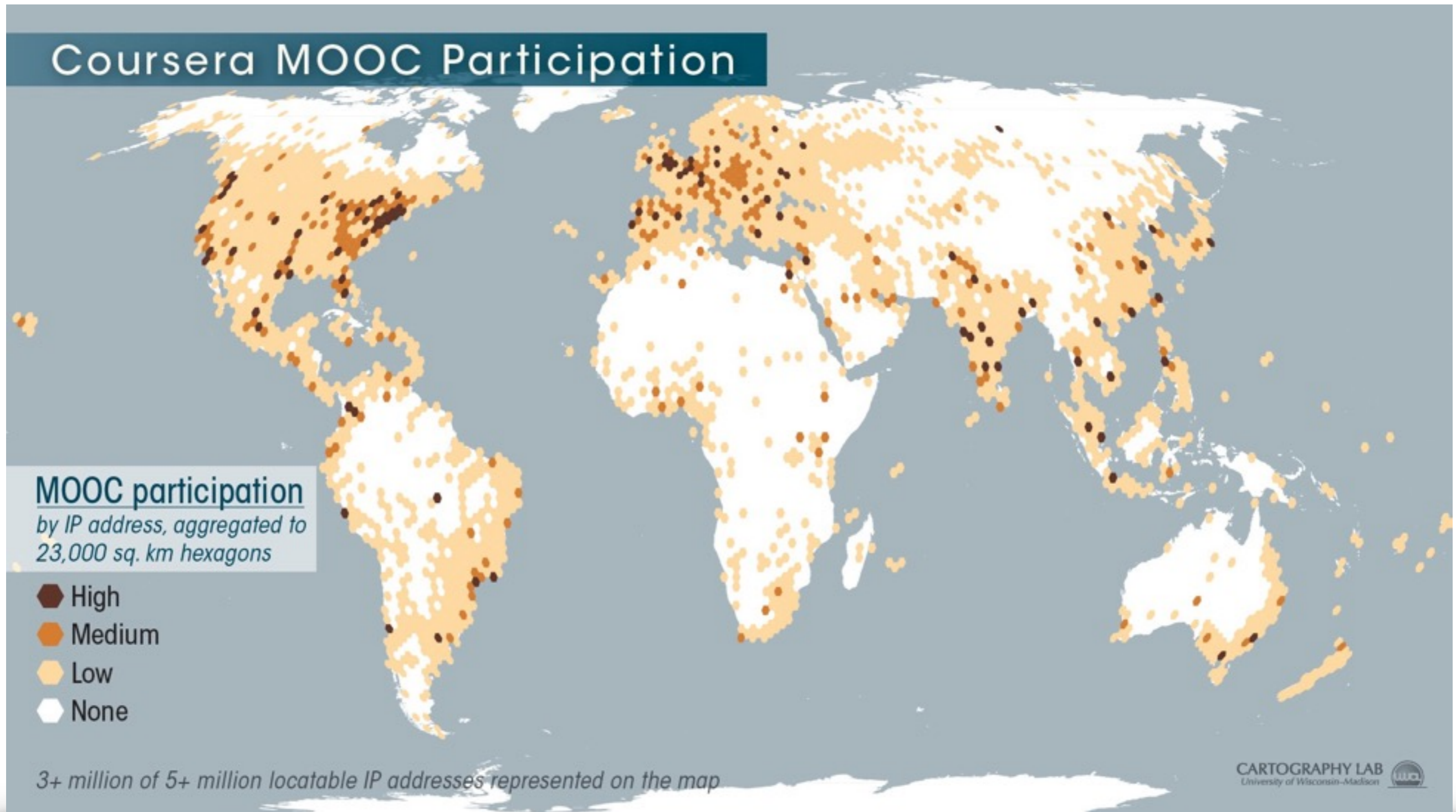
The output of the code is shown as:

```
*12345
*12345
*12345
```

The video player interface includes a progress bar at the bottom showing 0:00 / 5:14, a speed control set to 1.0x, and a volume icon. A large red play button is overlaid on the video content.



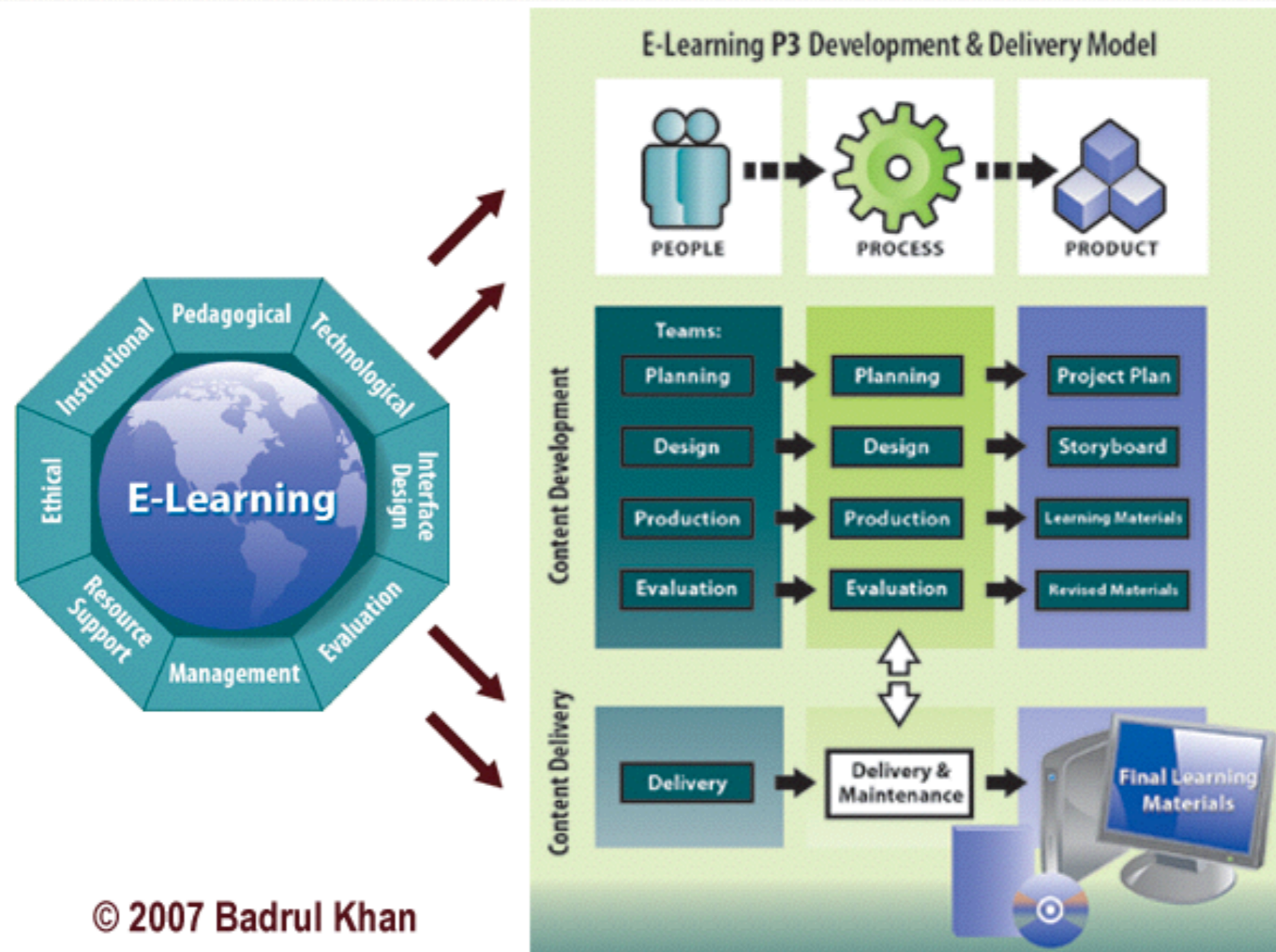
Peer Learning



Assessment Methods

The CAPEODL Model

Comprehensive Approach to Program Evaluation in Open and Distributed Learning



© 2007 Badrul Khan



Gamification



Learning Analytics



BigScholar 2014
THE FIRST WWW WORKSHOP ON BIG SCHOLARLY DATA: TOWARDS THE WEB OF SCHOLARS
SEOUL, KOREA, APRIL 8, 2014

Home CFP Organizers Submission Registrat

CALL FOR PAPERS (TXT)

Researchers worldwide are currently producing more and more sc books, patents, etc. Such data are big data by nature. For exampl provides bibliographic information on major computer science jour indexes more than 2.3 million articles records containing title, pag Concurrently, scholars are associated with various academic activ congresses, peer review, and so on. Such scenarios have motivat context of big scholarly data on a global scale. It is imperative and towards the innovative generation of values from Big Scholarly Da demands a re-evaluation of existing techniques, such as data min analysis. Furthermore, there is a demand for novel ways of develp foster the analysis and interpretation of social environments such

In this workshop, we will explore promising areas of research in bi emerging field of the Web of Scholars. This workshop also seeks i as:

- How to connect scholars on the web?
- How to facilitate collaboration among scholars?
- How to find the experts in a particular field?

Researchers are welcome to submit their papers that address the which may include, but are not limited to:

- Academic social network analysis
- Scientific recommendation
- Methods and tools for analyzing big scholarly data
- Indexing, searching, and mining scholarly data
- Connecting scholars using a Web approach
- Platforms and services for the Web of Scholars
- Web tools and techniques for big scholarly data
- Paradigms to promote scientific collaboration
- Scientific trends prediction
- Applications, use cases, and evaluations of big scholarly data

IMPORTANT DATES
Paper Submission Deadline: Jan 14, 2014 Jan 28, 2014
Author Notification: Feb 4, 2014
Final Manuscript: Feb 12, 2014



BigScholar 2015
THE SECOND WWW WORKSHOP ON BIG SCHOLARLY DATA: TOWARDS THE WEB OF SCHOLARS
FLORENCE, ITALY, MAY 18, 2015



BigScholar 2016
The Third WWW Workshop on Big Scholarly Data: Towards the Web of Scholars
Montreal, Canada, April 2016

Home CFP Organizers Submission Registration Program Keynote Speakers



Program Keynote Speakers

Program Keynote Speakers



Welcome to
BigScholar 2016

The Third WWW Workshop on
Big Scholarly Data: Towards the Web of Scholars
<http://msclab.org/bigscholar/>

A workshop of WWW 2016 (The 25th International World Wide Web Conference)
Montreal, Canada, April 11-15, 2016

The BigScholar 2016 workshop aims at bringing together researchers and practitioners working on Big Scholarly Data to discuss what are emerging research issues and how to explore the Web of Scholars.

emerging research issues and how to explore the Web of Scholars. and practitioners working on Big Scholarly Data to discuss what are The BigScholar 2016 workshop aims at bringing together researchers

Montreal, Canada, April 11-15, 2016

Important Dates

Paper submissions due:
December 22, 2015
Notification of acceptance:
February 2, 2016
Camera ready version due:
February 8, 2016
Workshop date:
TBA

Workshop date:

February 8, 2016

February 2, 2016

December 22, 2015

Notification of acceptance:

Workshop date:

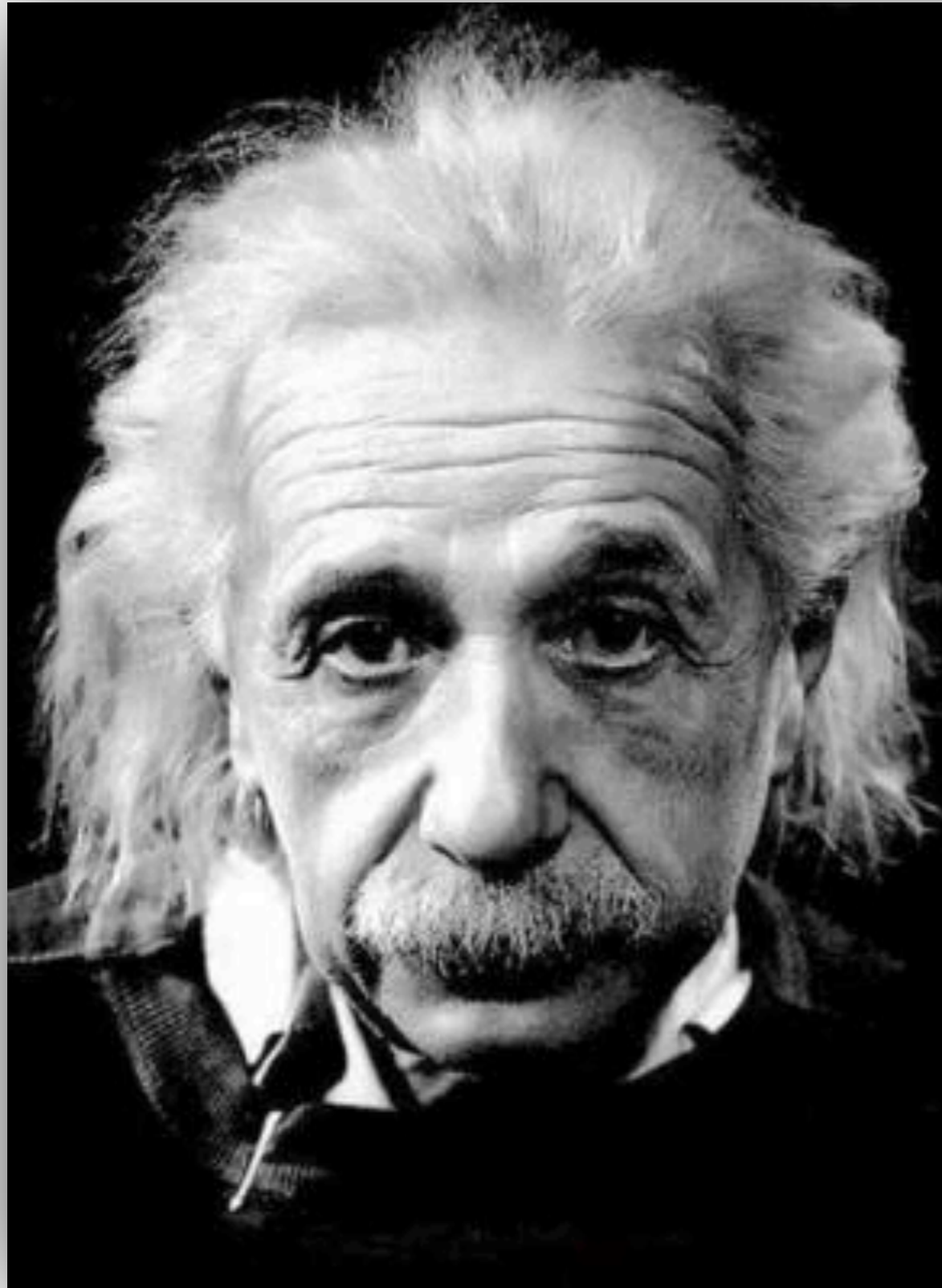
Big Education in the Era of Big Data by Irwin King @ IES/CSBio/IAIT/iNCEB 2015, Nov. 22-25, 2015, Bangkok, Thailand



Concluding Remarks

- *Be Inspired*
 - Big Education is the focus!
- *Be Informed*
 - Big Data in Education is the **VALUE** proposition!
- *Be Challenged*
 - Use technologies to transform education in the Big Data Era!





Once you
stop learning,
you start
dying...

Albert Einstein



An aerial photograph of a university campus, likely the Chinese University of Hong Kong, featuring a large green hillside with several white academic buildings, a prominent tower, and a satellite dish. The campus is surrounded by a blue bay and distant mountains under a clear sky. The text 'KEEP' is overlaid in large red letters at the top.

KEEP

Knowledge and **E**ducation
Exchange **P**latform

Acknowledgments

- Ken Chan (Ph.D.)
- Xixian Chen (Ph.D.)
- Chen Cheng (Ph.D.)
- Yuxin Su (Ph.D.)
- Haiqin Yang (Postdoc)
- Xiaotian Yu (Ph.D.)
- Connie Yuen (Ph.D.)
- Jichuan Zeng (Ph.D.)
- Hongyi Zhang (Ph.D.)
- Shenglin Zhao (Ph.D.)
- Jiani Zhang (Ph.D.)
- Tong Zhao (Ph.D.)
- Looking for more PhD students working on machine learning, Big Data, social computing, ...



Acknowledgments

- Brian Lau
- Patrick Lau
- Jesse Lee
- William Leung
- Cathy Li
- Lin Tsang
- Albert Yang
- Raymond Yuen
- Owen Zhang

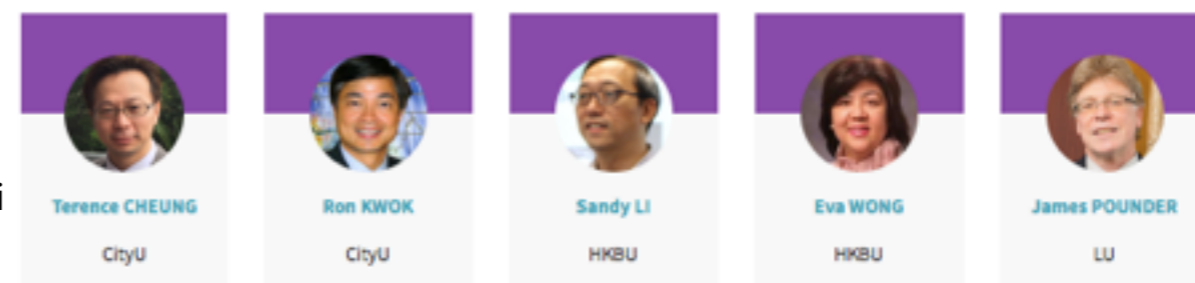
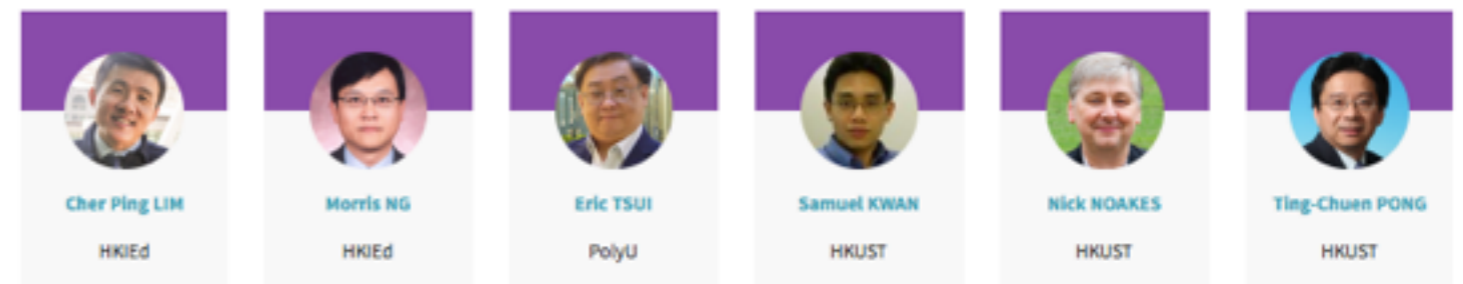
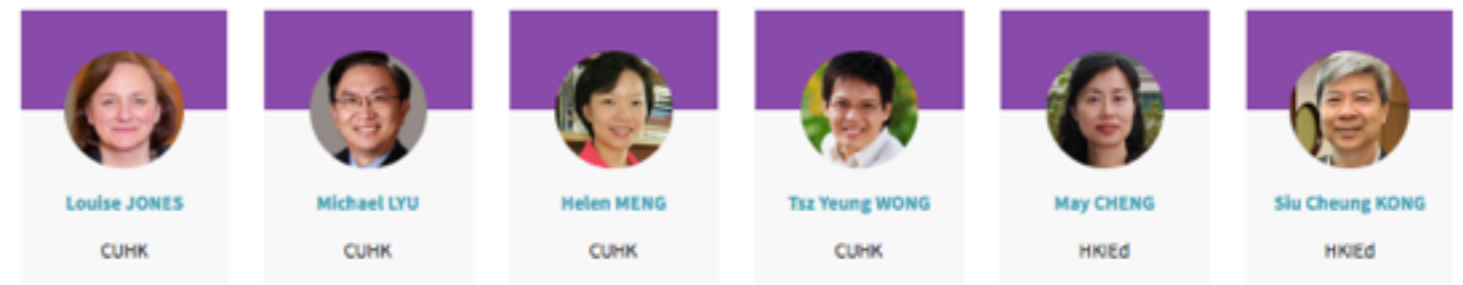
Principal Investigators



Advisory Board



Co-investigators



Big Education in the Era of Big Data by Irwin Ki



CUHK Excellence

- The only university in Hong Kong having Nobel Laureates as faculty with five Distinguished Professors-at-Large



Prof. Yang
Chen-Ning,
Nobel
Laureate in
Physics



Prof. Charles
Kao
Nobel
Laureate in
Physics



Prof. Sir James
A. Mirrlees,
Nobel Laureate
in Economic
Sciences



Prof. Yau
Shing-Tung,
Fields Medalist



Prof. Andrew Yao,
Turing Award

- **Nine** academicians of Chinese Academy of Sciences and Chinese Academy of Engineering





The Chinese University of Hong Kong



Q&A

